STUDY AND ANALYSIS OF STUDENTS ACADEMIC PERFORMANCE USING ARTIFICIAL INTELLIGENCE

A Thesis Submitted
in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY
in
Computer Science and Engineering

by
Rahul
(2K18/PHD/CO/19)

Under the Supervision of
Prof. Rahul Katarya
Department of Computer Science and Engineering
Delhi Technological University

Department of Computer Science and Engineering

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

August, 2024

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I Rahul, hereby certify that the work which is being presented in the thesis entitled "Study And Analysis Of Students Academic Performance Using Artificial Intelligence" in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy, submitted in the Department of Computer Science and Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from 2018 to 2024 under the supervision of Prof. Rahul Katarya.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor **Signature of External Examiner**

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR(s)

Certified that **Rahul** (2K18/PHD/CO/19) has carried out their search work presented in this thesis entitled **"Study And Analysis Of Students Academic Performance Using Artificial Intelligence"** for the award of **Doctor of Philosophy** from Department of Computer Science and Engineering, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Prof. Rahul Katarya
Supervisor
Department of CSE
Delhi Technological University

Place: New Delhi
Date: 14 August, 2024

Study And Analysis Of Students Academic Performance Using Artificial Intelligence

Rahul

ABSTRACT

This thesis explores multiple innovative machine learning and deep learning approaches to accurately predict student performance, providing educational institutions with valuable insights for improving academic fineness and identifying students at high risk for academic challenges. The research introduces four distinct models: Transient Search Capsule Network dependent on Deep Autoencoder (TSCNDE), Gannet Hunt Long Short Term Memory (Gannet Hunt-LSTM), Discriminability-Enhanced Transformer Architecture, and Explainable Deep Learning based Knowledge Distillation Framework (EDL_KDF). Each approach aims to enhance prediction precision, recall, accuracy, and other performance metrics, contributing to the understanding of how students' academic achievement can be forecasted more effectively.

Initially, the TSCNDE model uses a novel approach combining a deep autoencoder and transient search capsule network to anticipate the execution of student. The model demonstrated high accuracy and precision by processing data from Open University Learning Analytics Dataset (OULAD), with to the execution metrics achieving up to 99.2% accuracy and 99.8% precision. By focusing on crucial features and the influence of students' online activity, the method provides a reliable tool for assessing student success.Furthermore, the Gannet Hunt-LSTM model integrates a unique hybrid optimization approach by combining the behavior of the Gannet and Harris Hawk for optimal feature selection and classifier learning. This method minimizes overfitting issues and reduces information loss, resulting in exceptional prediction metrics such as 99.9% F-measure, 99.5738% recall, and 99.99% precision. The model leverages deep learning and LSTM networks to understand effect of students' social networking services utilization on academic performance.

Additionally, the Discriminability-Enhanced Transformer Architecture aims to improve feature-based prediction by using a hybrid DL approach to analyze students' academic performance. The model utilizes distance- and density-based outlier detection and multi-scale entropy techniques to enhance the prediction process. By extracting features and eliminating data inconsistencies, the approach excels in comparison to existing models and supports educational institutions in better understanding student learning behaviors. Finally, the EDL_KDF model leverages a deep explainable distillation approach with SHapley Additive Explanations (SHAP) to estimate endangered students as well as provide tactics for early intervention. By combining knowledge distillation framework with improved data augmentation techniques, the model achieves 99.6% accuracy and 99.21% precision over the OULAD dataset. The work supports educational institutions' efforts to improve teaching quality, enhance students' academic success, and optimize Virtual Learning Environments (VLEs).

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

ACKNOWLEDGEMENT

I take the opportunity to express my sincere gratitude to **Prof. Rahul Katarya**, Department of Computer Science and Engineering, Delhi Technological University, Delhi, for providing valuable guidance and constant encouragement throughout the work. His knowledge and expertise helped me at every stage of preparation of this research plan.

I would like to take this opportunity to thank all the faculty members and staff members of the Computer Science & Engineering Department, Delhi Technological University, for their encouragement and support.

I feel fortunate to have a spouse, Mrs. Monika and daughter Ivanka Chandra, whose unbounded support kept my morale up and gave me positive reviews. I would like to extend deep gratitude to have a great family who has supported me throughout the journey of my research work.

Last but not least, without whom I could not imagine being enrolled in a Ph.D., my father, Dr. Satish Chandra, and my mother, Smt. Rajni, for their unwavering encouragement, undeterred faith in me, and being the biggest pillar of strength, which supported me till the end. It would not have been possible to complete my research without their love, encouragement, and support.

Rahul

2K18/PHDCO/19

Ph.D. Scholar

Department of CSE

Delhi Technological University

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABBREVATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Networks
BiLSTM	Bidirectional Long Short Term Memory
CFP	Contextual Feature Perception
CNN	Convolutional Neural Network
DBNs	Deep Belief Networks
DE	Discriminability Enhancement
DEEDS	Digital Electronics Education and Design Suite
DL	Deep Learning
DM	Data Mining
DNN	Deep Neural Network
EDL_KDF	Explainable Deep Learning based Knowledge Distillation Framework
ES	Stacking Classifier
FEO	Fuzzy Equilibrium Optimizer
FN	False Negatives
GHO	Gannet Hunt Optimization
GPA	Grade point average
ISMOTE	Improved Synthetic Minority Oversampling Technique
KD	Knowledge Distillation
KNN	K-Nearest Neighbors
LA	Learning Analytics
LFE	Local Feature Extraction
LFF	Local Feature Fusion
LGOD	Local Gravity-based Outlier Detection
LMS	Learning Management Systems

LOF	Local Outlier Factor
LSTM	Long Short-Term Memory
ML	Machine Learning
MOD	Mean-shift Outlier Detection
MSE	Mean square Error
MSE	Multi-scale Entropy
NEIS	National Education Information System
OULAD	Open University Learning Analytics Dataset
PFA	Performance Factors Analysis
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristics
SHAP	SHapley Additive ExPlanations
SL	Student Learning
SVMs	Support Vector Machine
TEL	Technology-enhanced learning
TP	True Positives
TSCNDE	Transient Search Capsule Network-based Deep Auto Encoder
TSCNDE	Transient Search Capsule Network Based on Deep Auto Encoder
UPSA	University of Professional Studies Accra

CHAPTER 1

INTRODUCTION

This chapter presents an introduction on predicting performance of students and factors associated with student’s performance analysis and factors associated with it. This chapter includes problem statement, objective, motivation, contribution, thesis organization and finally concluded with a summary.

1.1 OVERVIEW

As the most effective instrument for fostering national development, education is widely acknowledged as the cornerstone of progress in contemporary society. But the educational environment has changed drastically, becoming more like corporate systems where organizations fiercely compete with one another for growth and revenue. In all of this change, the core goal has remained constant: giving kids a top-notch education so they may have successful futures [1]. However, contemporary institutions have several difficulties, especially when it comes to measuring and improving student productivity. Aware of the critical role that education plays in influencing people's lives, academics and educators are steadly anxious habits to progress student execution and foster a learning atmosphere [2]. Accurately predicting student performance is essential for educational institutions to improve teaching quality, recognize pupils who could be in danger of falling behind, and implement timely interventions. However, traditional methodologies encounter difficulties due to the extensive volume of data gathered and limitations in achieving precise performance predictions [3]. Predictive analytics becomes more and more necessary as educational institutions confront a variety of issues, such as improving curriculum and student retention [4].

1.2 NEED OF PREDICTING STUDENTS’ PERFORMANCE

Predicting the intellectual execution of students is a critical component in educational environments such as colleges and universities. The predictive analysis is an essential tool for developing strategies meant to increase academic performance and decrease problems like dropout rates. Through the utilization of predictive models, educational institutions may promptly identify students who exhibit signs of underperformance or disengagement. These interventions might be focused on addressing particular issues that each student is facing, customized learning strategies, or support networks that are specifically designed for them. In the end, the goal is to establish a setting that supports academic achievement, encourages student

retention, and optimizes each learner's potential. Thus, predictive analytics in education serves as a proactive tool in fostering student achievement and ensuring their continued engagement in the learning process [5].
Accurately predicting student performance has a big influence on a lot of people, such students, instructors, and educational institutions. Accurate and explainable forecasts are equally crucial in this area. Here, accuracy is defined as correctness of the predicted value, as well as explainability as predictor's capability to be understood [6].

The success of an educational institution is determined by a number of elements, one of which is the performance of the students. As a result, the institution is easily recognized by both students and various professional bodies. A number of factors are used to measure an educational institution's performance; the institution must enhance its ranking based on these characteristics. Educational institutions prioritize delivering high-quality instruction in order to get exceptional results. One of the significant metrics for better schooling institutions is performance of its students. As a result, to increase student performance and achieve intended results, institutions must monitor how their students are doing. Predicting student performance might be a potential subject to investigate in order to increase student performance, since it will give professional bodies and students advance knowledge about their future performance. A number of factors are used to measure an educational institution's performance; the institution must enhance its ranking based on these characteristics. Educational institutions prioritize delivering high-quality instruction in order to get exceptional results. One of the most important metrics for higher education institutions is the performance of its students. Institutions can monitor students' future performance and outcomes with the use of prior knowledge or information about their performance patterns. This has been the direction of the current study.

The prediction models created and examined in this study will assist educational institutions in making judgments at the start of each semester on how best to support slow learners, or underachievers, as a result. To obtain a more accurate assessment of the students' performance going forward, it was necessary to examine the admitted students' prior academic performance in many exams along with their personal background data. Imply that a student's achievement in an academic program is dependent on some characteristics of their behavior. Students' behavior and surroundings have an impact on their performance. These characteristics are thought to have the potential to directly or indirectly affect their performance. These characteristics include their academic success, study spaces provided, and personal background data. These are the pupils' latent qualities that they must develop in order to compete with high achievers. The performance of engineering students may be predicted using predictive models, which will reveal information about slow learners. Teachers can help these slow learners do better in the future by providing support. Tutors also be intelligent to estimate which characteristics of their scholars are most important to their achievement. A few more qualities have been added to the list of attributes that have been chosen from the literature for this study. Variables, features, and characteristics are used interchangeably here.

1.3 ACADEMIC PERFORMANCE

Academic performance is sum of a student's study habits and capability to manage and complete various assignments assigned by their professors. Academic performance, or how effectively a student satisfies requirements set forth by the institution and the local government, is thought of as the success assessed by the educational instructions. Stated differently, academic performance refers to a person's capacity for factual recall and study as well as their ability to express their knowledge both orally and in writing. Academic performance, also known as academic achievement, refers to the extent to which a learner, mentor, or faculty has met their instructive objectives and may be quantified by education [7].

1.1.1 Factors responsible for Academic achievements

Rate of failure has an influence on society and the lives of scholars in a number of ways. It is the responsibility of educators and educational establishments to make significant strides toward lowering the proportion of rejected applicants. There are a plethora of elements that contribute to kids' academic achievement. IQ, drive, study habits, socioeconomic level, personality fit, and a plethora of other factors are a few among them. The way these elements interact determines academic success [8]. To find out which elements are more important in increasing academic progress, a number of studies in this area need to be done. Here are some examples of these factors.

- ***Self-concept*:** Human behavior has often been explained by the socio-psychological aspect of self-concept, which is formed in entities as an outcome of their contact with social media platforms in peer groups, families, and other social organizations. A person's self-concept is the most comprehensive self-description they can have at any one moment. Thus, the child's self-concept is something that grows via many learning experiences as he learns himself, as he integrates with people and his surroundings, and as he realizes what he can and cannot accomplish. The secret of self-confidence is one's self-concept. A person's self-confidence, the things he chooses to do, his relationships with others, and his overall success may all be significantly impacted by a lack of confidence.

- ***Study habits:*** "Study habit" is one of the most crucial elements influencing academic performance. Studying requires a lot of work in order to get knowledge. One way to think about study is as a structured approach to topic mastery. Its main goals are to: (i) gain habits and information that will help one analyze concepts and deal with new situations; (ii) hone abilities; and (iii) form attitudes. Each of these definitions of study habits emphasizes a different facet of study strategies for efficient learning, such as established goals, mandated behaviors, exercising self-control, focusing efforts on achieving the correct goal, etc.

- ***Personality adjustment*****:** The process of constant contact between an organism and its surroundings is called adjustment. From the time of their birth, children are subjected to a multitude of societal agencies, including their family, school, religious institutions, peer groups, clubs, and numerous cultural organizations. These behaviors, the social environment in which the kid is raised, and the way the youngster responds to and handles these circumstances all have a role in how the child's personality develops. It is also referred to as a "Process involving better mental and behavioral responses by which an individual strives to cope successfully with inner needs, tensions, frustrations and conflicts and to effect a degree of harmony between the inner demands and those imposed on him by the objective world in which he lives." Personality maladjustment has a negative impact on learning, thinking, concentration, and other mental processes. Consequently, the person performs below his or her intellectual capacity; on the other hand, personality adjustment may lead to a better degree of academic performance. Put another way, a student who is well-adjusted is able to deal with the harsh realities of life and is brave enough to confront any challenge that stands in the way of his academic goals in an honest and detached manner. As a result, his academic achievement may be higher than that of a student who is not well-adjusted.

- ***Educational aspiration*****:** Aspiration is defined as the ability to set objectives. It is also claimed to be a person's source of self-motivation. In essence, aspiration is the goal-seeking conduct of a person who expresses worry for his future based on his current circumstances. An individual's aspiration level is defined as a quantifiable indicator of his potential for success in any given activity in the future. The ability to set objectives is referred to as aspiration# It is also believed to be a person's internal driving force. A person who expresses concern for their future based on their current circumstances is exhibiting aspiration. An individual's degree of aspiration is defined as a quantifiable indicator of his expected future achievement in any given activity.

- ***Manifest anxiety*:** Anxiety is becoming more and more acknowledged as a significant influence on modern life, and anxiety phenomena are currently a topic of interest in literature, science, the arts, and other areas of society. Anxiety levels that are manageable for the person are beneficial for learning in both adults and children. However, a huge "dose" might counteract the intended effect by triggering an overreaction or full paralysis, just like any other stimulus. Excessive anxiety might lead to abnormal and disjunctive behavior. Stated differently, anxiety is defined as a state of elevated and agitating internal strain, coupled with an ephemeral and frequently unsettling sense of unease and worry. Since anxiety heightens drive states, it may serve as a source of inspiration for bringing life to the classroom. While some learning seems to be aided by mild anxiety, learning appears to be hindered by extreme and low anxiety. Compared to the high anxiety group, children with low anxiety levels consistently completed the activity with fewer mistakes. For schoolwork, moderate anxiety might be preferable to extremely low or extremely high anxiety. On the other hand, anxiety appears to be necessary and beneficial for life to some extent. As most experts have emphasized, moderate anxiety is a natural part of life. It is necessary because it encourages individuals to take on their issues head-on. Without incentives, pressures, and a small amount of irritation, individuals wouldn't be driven to use them. Generally speaking, normal anxiety is felt in reaction to an impending catastrophe and goes away when the crisis has passed. When necessary, a person with moderate anxiety is prepared to act quickly and forcefully. Compared to a calm individual, a person who is somewhat nervous is watchful, cautious, and more likely to respond negatively to even the smallest amount of stimuli.

- ***Socio-economic status:*** The family is the first social unit in existence; it is responsible for both socializing the kid and providing for their bodily necessities and wants for affection. It is a collection of dynamic characters. Every member of this unit plays a certain role and has a certain status in relation to society; these statuses are based on factors such as financial standing, degree of education and cultural achievement, political standing, and so forth. The Learning Committee noted that a youngster from an upper class family with parents who have the greatest level of education does not have the same opportunities as a youngster from a rural slum with uneducated parents. Three factors make up one's socioeconomic status: income, occupation, and education. It is anticipated that a parent's education will have an impact on their children's academic success. The type of facilities offered for the children's education may depend on the parents' income. The parents' occupation determines the family's income and social standing. It is assumed that parents with high social status join reputable educational institutions. This will encourage the family's youngster to be well-equipped with more conveniences and knowledgeable about their environment.

- ***Intelligence:*** One of the main variables impacting academic success is intelligence. Studies reveal that students' More than any other quality, intellect is the most strongly correlated with their scholastic achievement. In schools and universities, students with higher intellectual capacity receive better grades than those with lower intelligence. The Greek phrase employed by Aristotle to refer to all cognitive processes is translated into Latin by Cicero, who also invented the term intelligence. It was believed that human nature innately has this ability to reason. Every definition discusses a different facet of intelligence. These various definitions, in a sense, emphasize three different facets of intelligent conduct. The capacities of learning, continuing abstract thought, and environment adaptation.

1.4 SOCIAL MEDIA

Social Media (SM) refers to computer-mediated technologies that let students produce, share, and trade knowledge, images, concepts, and videos with other students and virtual communities. For a number of years, use of SM in classroom has generated debate. The potential consequences of SM use in the classroom have scared off a lot of parents and teachers. Social media usage is expanding quickly on a global scale. A growing number of adults and teens are interacting with friends, relatives, and strangers on websites like Facebook, MySpace, Skype, WhatsApp, and X. Knowledge workers may broaden the breadth of their professional relationships by using social networking sites to help the community find expertise, share material, and work organized to make satisfied. By employing online community, investigators can have access to a network of persons and info that covers beyond their traditional social circle. ResearchGate is one social networking site created especially for academics. By using social networking sites, educators can promote reflective analysis and the growth of an online learning community. A plethora of information was available on a teacher's Facebook page, which also showed positive attitudes towards the instructor and the course and predicted improved motivation and emotional learning. Social networking gives teachers a fantastic opportunity to forge their own professional identities and expand their proficient network by facilitating idea exchanges and debate participation. Teachers may find valuable knowledge on social networking sites that will increase their ability to handle particular circumstances. Also, students could feel more at ease contacting teacher educators who are approachable and in person or who converse with them informally on Facebook, WhatsApp, and ResearchGate; this provides them with the motivation they require [9].

1.1.2 Influence of social media in education

Knowledge is created and shared through SM networking, and these functions are extremely valuable when it comes to higher education. SM is a big part of students' lives and the area of education. Social media makes it easier and easier to interact, share information, and obtain information. These social media platforms allow teachers and students to collaborate and support one another while furthering their education. Teachers utilize social media to impart knowledge by setting up profiles and groups where students may access the material. Teachers may collaborate on ideas and refer their pupils to Facebook, LinkedIn, WhatsApp, and Skype. Teachers design hashtags that let students annotate their academic postings and browse contributions to see what the group has come up with artistically. Teachers may use social media to teach their pupils, which is one of the key reasons they have adapted to using it in the classroom. Not only do they manage to simplify the job, but they are also growing professionally and making a name for themselves in the community. Professors may be seen excelling in their work on Facebook, X, blogs, and YouTube, to name a few instances.

Because these social media sites are widely used, they can aid academics in building a solid reputation in their field. WhatsApp, blogs, and Wikis are better for schooling and learning process; Facebook and LinkedIn are better for networking and establishing professional and social ties. When it comes to leveraging social media to engage audiences, departments such as Admissions, Enrollment Management, Public Relations, and Student Services are missing a crucial component social media for community building. Every college and university is utilizing SM to the fullest extent feasible in the classroom for a variety of subjects, including admissions, campus life, and alumni connections. SM is a part of everyday life for both teachers and students. Academicians risk alienating a large number of prospective audience members if they ignore the usage of SM. In higher education institutions, its application can show to be a highly efficacious measure [10].

1.1.2.1 Social media in teaching and learning

SM platforms like Facebook, X, Google Plus, Flickr, and open social practices like blogging are being utilized in education to facilitate easy communication between students in peer groups and possibly with individuals outside the classroom, like experts in the field and other students studying related subjects. Due to fact that these SM platforms are widely accessible, it is necessary to carefully weigh the dangers associated with transparency and maintain constant contact with students for addressing academic concerns and handle problems related to their usage of SM as they emerge. The scholarly advantages of candid conversation and scholarly debate in real-world online settings outweigh these disadvantages. The popularity of a new social networking platform that focuses on fostering relationships with kids outside of the classroom is rising. There are several benefits to utilizing SM for educational purposes. Based on a research, SM employes increases learning chances for students, facilitates real-time connection outside of the classroom, encourages collaborative opportunities, and boosts creativity. In order to continue their discussion with teacher’s online, students can watch educationally relevant films or share information about what they have seen and learnt. Through their connections on social media, instructors may also get knowledge from their pupils. In a similar vein, a teacher can keep an eye on their pupils as they absorb knowledge, think critically, share, engage, and summarize conversations. SM provides users a way to stay in touch with their professors and classmates no matter where they are, giving them the flexibility of longer work hours. Certain features of SM, particularly those on Facebook, WhatsApp, YouTube, and the Kaizala app, may encourage students to participate in innovative and social learning processes that take place outside of conventional classrooms and educational institutions [11].

1.1.2.2 Social media for academic purpose

Utilization of social media by university students is undoubtedly an interesting area of study for social scientists and educationists. Remind them that there are advantageous designs and methods for implementing it at school level in the literature that is now available. It explains the contents' introduction and concentrates on sharing, interacting, collaborating, and socializing that may happen through its use. Various compelling arguments seem to support the utilization of SM in educational institutions. Facilitating communication among and between students in virtual communities is a common practice. Facebook was recommended as a communication tool for contacting students, among other things, and it seems to be the most popular one. Internet sites place a lot of emphasis on creating virtual cities that correspond with shared hobbies or pastimes. Give them the instruments necessary to enable them to accomplish this. In the context of collaborative learning, online resources addressed the significant issue raised by educators in the last few months. For example, deficiencies in the humanitarian component and acknowledging that its instruction lacks soul. However, websites have helped some people find solutions to their issues through interactive affiliates by connecting them with others, making human engagement in the academic process something significant. This was introduced along with a rise in the popularity of collaborative learning, and several social science researchers conducted studies to examine this phenomena and provide an explanation for why people are drawn to websites, for instance.

Additionally, teachers have noted that utilizing online resources can promote online communication between students outside of the classroom and outside of the typical classroom context [12]. When students make changes to their individual webpages on websites or blogs and update their user accounts, they have the option to include extensive personal information about themselves, such as their full name, date of birth, educational background, address, and hobbies. By looking into the students' profiles, teachers who are utilizing these technologies in the classroom will be able to learn a great deal more about the pupils they are instructing [13].As a result, educators are able to customize this curriculum by utilizing the individual student profiles [14]. Advanced schooling students are refreshed to involve in investigative-based and collaborative events using certain social media platforms, like Wikipedia and blogs. Opportunities for successful learning may arise as a result of this wide access to active engagement. IN keeping with the idea of students developing collaboratively, social media gives students and teachers the opportunity to publish and exchange information that comes from learning activities (such as course materials like test cases, assignments, notes, and syllabus) and get peer feedback. Publication and presentation of their work to a broad audience via blogs, wikis, or podcasts, learners get chance to adapt new concepts and modify their own knowledge via introspection [15].

Students can work cooperatively, especially those at higher learning levels, by investigating the options provided by the online social environment to address specific academic problems or problems with their classmates [16].This suggests that social media integration might facilitate collaborative or team learning, allowing students to build good relationships while working toward specific end goals in both offline and online modalities [17]., a social networking platform for college students was developed with the goal of enhancing social interaction and collaborative learning. According to their research, including social networking technologies in conventional education can draw students and even encourage their engagement in the process. In other comparable research, wiki software, blogging services, and social bookmark posting tools are well-versed in getting students involved in group projects and promoting the creation, modification, and discussion of information. Because social networking learning fosters the development of communal knowledge and active user engagement, the study guarantees that it is promising to take into consideration [18].

1.1.2.3 Positive effects of Social Network Site Usage among Adolescents.

- ***Socialization:*** Developing a social network and making friends are crucial aspects of growing up. Why is this relevant? The reason for this is because it enables the kids to share ideas and pick up new skills. They learn more the more they engage in. They will undoubtedly grow more self-assured as a result of this.

- ***Knowledge sharing:*** Provides pupils a simple and efficient approach to exchange knowledge. Everyone has access to online data, excepting limits imposed by those who share it. Pupils may quickly access the information, study it, make changes as needed, and share it. As a result, knowledge flows more easily.

- ***Updating oneself:*** Without social networking technology, how else might students learn about global events? They'll be able to learn new information as it becomes available. They are able to upgrade their own knowledge base as a result.

- ***Learning from several sources:*** Students may select the organization, activity, or individual they wish to follow for daily updates thanks to the way social networking tools are designed. Social networking tools enable students to follow educational websites, educators, potential higher education institutions, technology they wish to stay up to date on, and much more. There are several methods of learning.

- ***Being ready:*** The key feature of social networking technologies is their ability to display global trends. By monitoring the updates, sharing, discussions, and conversations that take place on social networking sites, students might gain insight into what could be expected of them in the future.

- ***Communicating their ideas and feelings:*** Students sometimes lack the appropriate forum to express their feelings and opinions. They have the ability to voice their opinions thanks to social networking tools. Students find it easier to distinguish between right and wrong when they are given the freedom to express their thoughts and feelings. They usually get more adept at making judgments.

- ***Creativity development:*** More creative works are being shared as a result of how quick and simple it is for people to post images, videos, or narratives. Students who are able to receive prompt feedback on their artistic endeavors from friends and family are able to hone and improve their artistic skills, gain much-needed confidence, and make career decisions.

- ***College Entrance:*** Students may obtain up-to-date information from various social media platforms on the admission procedure, schedule, and other prerequisites [19].

1.1.2.4 Negative effects of Social Network Site Usage among Adolescents

- ***Lack of face-to-face communication:*** Students no longer have capacity to participate in face-to-face conversation because of social media.

- ***Time Wastage:*** Students become drawn to social networking sites when they are researching and studying online, and occasionally they lose sight of the purpose of utilizing the internet. Students' time is wasted, and occasionally they are unable to turn in their work by the deadline.

- ***Loss of Motivation:*** Use of these social networking sites lowers motivating level of the students. Rather than acquiring real-world experience and practical knowledge, they depend on the virtual environment [20].

- ***Cyberbullying:*** A study released by PewCenter.org claims that the majority of kids have experienced cyberbullying in the past. Bullying online has become quite simple as anyone can make a fictitious account and do anything without being discovered. Rumors, intimidation messages, and threats can be disseminated to the general public to incite unrest and disorder in society.

- ***Hacking:*** It's simple to hack someone's personal information and distribute it online. It may result in losses in terms of money and personal life. In a similar vein, identity theft is another problem that can cause financial losses for anybody through access to compromised personal information. In the past, someone has hacked into a number of people's personal Facebook and X accounts, posting things that have impacted the victims' personal life.

- ***Addiction:*** Social media's addictive qualities are terrible and can interfere with personal relationships. The majority of people impacted by SM addiction are teens. They develop a strong commitment and are eventually discarded by society. Additionally, it might squander personal time that could be better spent on worthwhile endeavors.

- ***Fraud & Scams:*** There are several examples of people defrauding others and committing fraud utilising SM. For instance, the five most common social media frauds are included in this list.

- ***Security Issues:*** Security agencies nowadays have entree to people's personal accounts. It practically compromises privacy. You never know when an investigator may come see you about anything you may have inadvertently or unintentionally discussed online.

- ***Reputation:*** A bogus tale may quickly damage someone's reputation on social media by propagating it widely. In a similar vein, companies may potentially lose money as a result of negative information spreading over social media.

- ***Health Problems:*** Excessive use of social media might be harmful to one's health. Since exercise is the key to losing weight, most individuals get sluggish as a result of using social networking sites excessively, which disrupts their daily routines. This study from Discovery will astound you with how negatively using SM may impact your health.

- ***Glamorizes Bad Habits:*** Drawback of SM is that users like to follow affluent or drug-addicted people and publish their opinions and videos online, which eventually encourages other users to do the same and develop drug and alcohol addictions.

- ***Virtual World:*** Most students squander time on social networking sites and stay in the virtual world. Although it may appear that a kid has thousands of social media pals, in actuality, the only people who can support him or her in a trying circumstance are his or her parents [19].

1.5 PROBLEM STATEMENT

The pervasive use of social media among the younger population is undeniably impacting academic performance, necessitating the development of methods to predict and guide students in utilizing technology effectively to achieve their academic goals. Existing student performance analysis methods, predominantly employing machine learning approaches, face several challenges. Firstly, these methods often require manual feature extraction prior to classification, making them labor-intensive and potentially subjective. Moreover, while machine learning techniques can excel with smaller datasets, their performance tends to degrade with larger datasets, and the selection of relevant attributes significantly influences accuracy. Inaccuracies may arise from considering irrelevant attributes, rendering detection results unreliable. Furthermore, issues such as instability in performance classification, lazy learning capabilities, inability to handle noisy data, and computational complexity all contribute to limiting the accuracy of student performance analysis. Despite the promise of Artificial Intelligence (AI)-based learning analytics in supporting educators by collecting, analyzing, and reporting learner data, predicting student performance in online platforms remains challenging. Complex inputs, over fitting problems, reduced learning ability, and inefficiencies in existing prediction models all contribute to the limitations of current approaches. Thus, there is a pressing need to address these challenges to develop more effective methods for finding out student performance accurately in virtual environments.

1.6 RESEARCH QUESTIONS

- How does the hybrid Transient Search Capsule Network-oriented Deep Auto Encoder (TSCNDE) model compare to traditional ML and DL methods in expecting student execution accurately in online environments?

- What are the key features extracted by the enhanced feature extraction mechanism, and how do they contribute to the improved performance of the classifier in student performance prediction?

- How does the Explainable Deep Learning based Knowledge Distillation Framework (EDL_KDF) reduce computational complexity while maintaining or improving prediction accuracy in student academic performance?

- What are the most significant challenges faced in handling large datasets for student performance prediction, and how does the novel outlier detection model address these challenges efficiently?

- How does the utilization of advanced statistical techniques and machine learning algorithms in the outlier detection model enhance the overall efficiency of the student performance prediction system, particularly in terms of processing time and accuracy?

1.7 MOTIVATION

Evaluating student performance is fundamental in guiding their educational journey towards achieving their full potential. However, traditional methods face significant challenges, particularly in handling the burgeoning volume of student data generated in today's educational landscape. As the learning process primarily unfolds within the intricate framework of a student's mind, understanding what students can do with their learning remains a complex endeavor. Academic performance serves as a tangible manifestation of this learning process, offering insights into students' comprehension, progress, and potential challenges. Yet, as datasets grow exponentially, conventional evaluation methods struggle to keep pace, hindering educators' ability to provide timely and targeted support to students. In response to these challenges, there is a pressing need to harness technological advancements to develop innovative solutions that can accurately predict and detect outliers in student performance data. By leveraging cutting-edge techniques such as DL and advanced outlier detection algorithms, this research targets to points the shortcomings of existing models and provide educators with a robust framework for early identification and intervention.

1.8 OBJECTIVES

- Develop and implement a hybrid TSCNDE aimed at improving student performance prediction by identifying strengths and weaknesses in academics.

- To utilize a novel DL model to accurately predict student performance, ensuring generation of precise results.

- To improve input quality by addressing missing values in the initial pre-processing stage, thereby streamlining subsequent stages of analysis.

- To introduce a novel Explainable Deep Learning based Knowledge Distillation Framework (EDL_KDF) to predict student academic performance with reduced computational complexity and heightened accuracy.

1.9 CONTRIBUTION

- Developed a hybrid TSCNDE model tailored specifically for student performance prediction. This innovative approach combines the strengths of capsule networks and deep autoencoders to capture intricate patterns in student data and identify both strong and weak academic areas.

- Evaluated the suggested TSCNDE method rigorously using diverse evaluation metrics to ensure its accuracy and reliability in predicting student performance. By employing a comprehensive set of evaluation criteria, including but not limited to precision, recall, F1-score, and accuracy, effectiveness of the model was thoroughly assessed.

- Correlated the performance of the TSCNDE model in contradiction of existing ML and DL methods commonly used for student performance prediction. Through extensive experimentation and analysis, demonstrated the advantage of the proposed approach in terms of accuracy and efficiency, thereby highlighting its potential to enhance educational outcomes.

- Established a novel outlier detection model designed to streamline the processing of large datasets. By leveraging advanced statistical techniques and machine learning algorithms, outlier detection model identifies anomalous data points efficiently, thereby reducing processing time and enhancing the whole efficiency of the prediction scheme.

- Implemented a cutting-edge deep learning model capable of accurately predicting student performance based on diverse input features. By harnessing the power of deep neural networks and leveraging sophisticated architectures, the model delivers precise and reliable predictions, enabling educators to tailor interventions and support systems effectively.

- Engineered an enhanced feature extraction mechanism to optimize the performance of the classifier. By carefully selecting and transforming relevant features from the dataset, the model's skill to discern meaningful patterns and make exact predictions is significantly improved, ensuring efficient operation and robust performance in real-world educational settings.

1.10 THESIS ORGANIZATION

1. Chapter 1: Introduction

Chapter 1 discusses about the Need of predicting students' performance, Academic Performance, Factors affecting academic performance, Social Media, Influence of social media in education, Negative impacts of SM and negative impacts of social media. Finally the chapter includes problem statement, objectives, motivation, contribution, thesis organization and concluded with a summary.

2. **Chapter 2: Literature Review**

   This chapter includes the relevant works based on Student Performance Prediction. The comparison table for corresponding reviewed methods are also given.

3. **Chapter 3: Deep Auto Encoder based on a Transient Search Capsule Network for Student Performance Prediction**

   Chapter include detailed description of data Pre-processing, Feature Extraction, Feature selection using Fuzzy Equilibrium Optimizer (FEO), future classification, Parameter Optimization using TSO algorithm and also discussed about the results of the proposed work.

4. **Chapter 4: Optimized Deep Learning Based Students Performance Analysis Based on the Influence of Social Media**

   This chapter discussed about the proposed hybrid optimized deep learning approach named Gannet Hunt Long Short Term Memory. Included the detailed description of data Acquisition, K-NN based Missing data Imputation, Feature Extraction, Gannet Hunt based Optimal Feature Selection, Gannet Hunt-LSTM for student performance prediction, LSTM for predicting the student performance and also discussed about the results and discussion.

5. **Chapter 5: Discriminability Enhanced Transformer Architecture for Students Performance Prediction Using Enhanced Features**

   This chapter discussed about the proposed Discriminability Enhanced Transformer Architecture for Students Performance Prediction Using Enhanced Features. This included the description of Distance based Method, Discriminability Enhanced Transformer Architecture and Result and Discussion.

6. **Chapter 6: Shapley Explainable Deep Learning based Knowledge Distillation Framework for Student's Performance Prediction**

   This chapter discussed about the proposed Explainable Deep Learning based Knowledge Distillation Framework (EDL_KDF). Pre-processing the input data using Missing value imputation, Data balance using Improved Synthetic Minority Oversampling Technique, Student Performance Prediction using Explainable Deep Learning based Knowledge Distillation Framework, Pre-trained Teacher Network using Dual Attention based Dense Bi-directional

LSTM Network etc. also included in this chapter. Finally results and discussion is also included.

7. **Chapter 7: Conclusion and future scope**

This chapter discussed about the conclusion of entire research and its directions for future enhancements.

1.11 SUMMARY

In contemporary society, education is widely recognized as the cornerstone of progress, yet the educational landscape has evolved into a competitive environment reminiscent of corporate systems. Despite these changes, the fundamental goal remains constant: providing students with a high-quality education for successful futures. However, educational institutions face several challenges, particularly in measuring and improving student productivity. Recognizing the serious role of tutoring in modelling lives, educators and academics continually seek ways to enhance student performance and create conducive learning environments. Predictive analytics has occurred as a vital tool in this pursuit, enabling institutions to accurately forecast student performance, improve teaching quality, identify at-risk students, and implement timely interventions. This predictive analysis is essential for developing strategies aimed at enhancing academic achievement and reducing issues like dropout rates.

CHAPTER 2

LITERATURE SURVEY

This chapter presents a literature survey on predict performance of students dependent on various parameters and examines the most effective techniques for analysing student performance prediction techniques. It develops research questions, points out gaps in the literature, and offers recommendations for further research based on the findings.

2.1 OVERVIEW

It is more crucial than ever to identify students' talents at early stages and direct them toward appropriate career pathways in an increasingly digital environment [21]. Nowadays colleges, universities, and institutions operate in very competitive and diverse world, and they are nearly equivalent to corporate organizations [22]. Colleges, education centers, universities, and other institutions accept students in many disciplines who meet certain grade requirements and have the potential to achieve every year. In the realm of education, Data Mining (DM) and other techniques are currently gaining popularity for use in performance research and teaching improvement [23]. The definition of DM is "a process that processes massive amounts of data using one or more software to find the masked patterns in large data lots" [24]. Education is currently using a great deal of DM and ML technologies to provide teachers and students achieve better teaching and learning outcomes [26]. To forecast students' success and ensure they receive a quality education, ML approaches and techniques can be applied to student data [25]. The major goal of this effort is to relate particular ML approaches to scholar data sets. This study provides a more effective analysis that would allow educators or administrators to separate pupils according to their performance levels in addition to predicting their failure [27]. The effectiveness of several models is compared by comparing the supervised classification algorithms. In order to improve their performance, remedial classes will be rescheduled for underprivileged students, and it will receive additional works. Universities will be able to develop guidelines for effective admissions with the assistance of this [28]. A number of technologies are being created, tested, and implemented into practice to address these security concerns [29]. Some of the related work based on this scenario is represented in this section.

2.2 REVIEW OF RELATED WORKS TO PREDICT PERFORMANCE OF STUDENTS DEPENDENT ON VARIOUS PARAMETERS

Numerous student performance prediction models are developed to predict outcome of students dependent on various parameters. In this case, examines the most effective techniques for analysing student performance prediction techniques. Some of the recent existing research studies carried out on student performance using different techniques are described below:

Amjad, S., *et al*. [30] state that machine learning criteria were used to assess academic achievement of high school learners while taking into account the influence of social media. This involves gathering characteristics from the dataset, such as parents' involvement, academic background, technical background, and demographics related to social media usage. The classifier derived from random forests had the best accuracy in identifying the student's performance. In this case, choosing the appropriate qualities improves the prediction's accuracy. Based on the data, it seems that students who were proficient in utilizing technology had performed better, but those who are not proficient in using it wasted time and effort.

Nti, I.K., *et al*. [31], had used machine learning driven academic performance prediction to analyze drawbacks of social media use. This study provided an actual database to predict academic achievement by taking into account variables such as exposure level, use frequency, kind, and length of social media use in classroom. Finally, ML classifier was used to predict academic success. The method's study shows that while students with low academic performance spent a lot of time on social media, good use of the platform had no effect on academic achievement.

Poudyal *et al*. [32], had suggested a hybrid 2-dimensional CNN (convolutional neural network) architecture for analyzing educational data. The model transforms one-dimensional data samples into two-dimensional image data, leveraging a hybrid deep learning algorithm. Initially, the dataset undergoes pre-processing via one-hot encoding to convert samples into numerical variables. Subsequently, a 2-dimensional representation of the data was generated using 37 numerical features, with zero padding applied and reshaped into a 40-length array to create an image-like structure. Two convolutional networks with diverse layer configurations were then employed to process the data and extract relevant features for analysis.

In order to extract statistical traits from a dataset used to predict student performance, Ghassen Ben Brahim [33] created a prediction model. Online session data, including metrics like text editing, the number of keystrokes, as well as the amount of time spent on different tasks, were gathered by the model. Combining activity type, time statistics, and peripheral activity yields 86 statistical characteristics. The entropy approach was then applied to these traits in order to pick them. Through classification, the effectiveness of the chosen features was assessed using a variety of techniques, counting as logistic regression, NBs, random forest (RF), SVMs, and multi-layer perceptrons.

Asselman *et al*. [34] utilized PFA strategy built on the Random Forest, AdaBoost, and XGBoost models in an effort to improve student performance prediction accuracy. It had tested the models that created on three distinct data. The experimental findings determined that, in comparison to original PFA algorithm, the scalable XGBoost significantly improved the performance prediction and outperformed the other assessed models. It used ensemble algorithms based on ML procedures to predict students' performance. Their study introduced a Performance Factors Analysis (PFA) method employing diverse models including AdaBoost, XGBoost, and RF. The developed model achieved an accuracy of 78.75% across three distinct datasets.

Kusumawardani *et al.* [35], had suggested a deep learning-based LA technique (transformer encoder) to forecast students' final performance in steps using log activity from an LMS. The goal was to identify students who were most possible to interventions can be implemented as soon as feasible. For daily or weekly prediction, suggested model was assessed by the Open University LA Dataset (OULAD). The findings validate that the model's early prediction accuracy on withdrawn versus pass-distinction classes was 83.17%. On the other hand, the accuracy was at least 76% in the early stages of the withdrawn flop versus pass-distinction and fail against pass-distinction tasks. The LSTM model and the suggested model were contrasted. Dependent on statistical testing, it was discovered that transformer encoder outperformed LSTM, with average difference values ranging from 1% to 3% in terms of accuracy and from 3% to 7% in terms of F1-score for all workloads. Additionally, an ablation research was carried out for the imbalanced class problem employing positional encoding, several feature aggregation techniques, and a weighted loss function. It discovered that positional encoding-free model performed better in every OULAD scenario. Additionally, in certain instances, weekly feature combination and utlize of a biased loss function yielded superior results.

A number of automated learning techniques were employed by Rivas *et al.* [36] to teach data to university students. Tree-based models and artificial neural networks, including decision trees, random forests, multi-layer perceptrons, hopfield networks, extreme gradient boosting simple perceptrons, competitive learning networks, CNN, and authorizing neural networks, were used in the series learning process. A publicly available dataset comprising information from 32593 students kept in a virtual learning environment was used to examine the system. The neural network uses the virtual learning environment to retrieve the dataset's instance for parametric research information. The methodology solely examined the pass, fail, and average classes in order to assess student performance.

Rodríguez-Hernández *et al*., [37], had introduced a methodology for employing ANN to assess academic performance of students in tertiary education. Their study encompassed a vast dataset comprising 162,030 student records from both public and private universities in Colombia, representing both male and female students. Through their implementation of the ANN model, the researchers achieved a noteworthy accuracy rate of 82%. When compared to other ML processes, artificial neural networks perform better in evaluation metrics like recall and F1 score. Additionally, it had been discovered that socioeconomic status, previous academic success, and high school attributes were significant indicators of students' academic success in postsecondary education. This study concludes with some suggestions for using artificial neural networks and some points to think about while analysing academic achievement in higher education.

Shahzadi, K., et al., [38] suggested a ML method for analyzing student performance by taking into account the association among online platform usage and academic achievement. For examination of student performance, a number of factors including school, social, demographic, and student grade were taken into account. According to the research, pupils who used social media and cell phones neglected their studies and performed poorly. SVM outperformed KNN in this analysis of the performance, which was based on accuracy and assessed using machine learning techniques. The results show an important relationship between students' academic achievement and social media involvement. This study demonstrates how most students had access to cell phones and the internet, which causes them to use it for 40 minutes a day for four hours a day without paying attention to their academics. These online events had an effect on students' academic execution. The backdrop of previous studies and an introduction will be given before this study begins. Following that, a research question based on background information will be presented.

Alshamaila, Y., *et al.* [39], had developed the student's performance prediction about social media's influence based on the intelligent method. For the purpose of predicting performance, a number of factors were taken into account, including amount of time spent on SM, usage of internet and accessibility of connections, the year of study, gender, age, and numerous other factors. When compared to the analysis conducted without data balancing, the analysis assessed by the intelligent technique with data balancing produced a superior result.

Waheed, Hajra, *et al.* [40], had employed a DNN to expect at-risk pupils and provides approaches for initial intrusion in situations based on group of special handmade elements composed from the clickstream data of virtual learning environments. According to findings, recommended approach can classify data with accuracy of 84%–93%. It demonstrate that standard logistic regression and SVM approaches were outperformed by a DNN. The accuracy of SVM was 79.95% - 89.14%, whilst LR reaches an accuracy of 79.82% - 85.60%. Consistent with previous study, the results show that adding legacy data and assessment-related data can greatly influence the model. It was seen that students who were eager to access the material from the earlier lectures perform better. The purpose of study was to provide higher education decision-makers create the framework that was required for pedagogical assistance in order to support sustainable education.

Giannakas, Filippos, *et al.* [41], had recommended a DNN approach for binary classification with two hidden layers in order to anticipate software engineering teams' performance in advance. Several optimizers (Adagrad and Adadelta) and activation functions were assess the framework. A dataset including more than 30,000 entries categorized into 74 groups was employed to train and assess the framework. Furthermore, approach was interpreted using SHapley Additive exPlanations (SHAP) technique, which allowed for the extraction of the most significant elements that either positively or negatively affected the final forecast. Among the other results, it was demonstrated that the framework's learning performance was 86.57% and 80.76%, indicating that it was capable of accurately and sufficiently predicting teams' performance. When Adadelta and Adagrad optimizers were used, the prediction accuracy was found to be 76.73% and 82.39%.

Tsiakmaki, Maria, *et al.* [42], had determined deep neural network transfer learning works for predicting students' achievement in advanced teaching. The significant advancement in the field of educational data mining since there hasn't been much investigation done on the creation of prediction models using transfer learning techniques. As a result, a large number of experiments were carried out using information from five required courses from two apprentice plans. The trial outcome show that, in most circumstances, it was possible to predict the prognosis of students who at danger of failing with a sufficient degree of accuracy, given the availability of datasets containing information on students who had taken similar courses.

Gray, Cameron C., and Dave Perkins. [43], had examined at the research on the techniques and metrics used in learning analytics today. Although there were methods available, it was found that they do not prioritize identifying difficult kids as soon as possible. In order to improve a predictive approach, this work employs contemporary ML tools and techniques along with the definition of a new descriptive statistic for student attendance. It shows how students can be recognized with about 97% accuracy as early as week 3 semester. In addition, place this finding in a suitable pedagogical framework to encourage its application as a component of a more all-encompassing system of student assistance.

Hussain, Mushtaq, *et al.* [44], had suggested to forecast disadvantages that students would face during a later session of the digital design course. Digital electronics education as well as design suite (DEEDS) was a technology-enhanced learning (TEL) system that data was examined by using ML methods. ANNs, SVMs, LR, DTs, and NB classifiers were among the ML algorithms. Students can log data while solving varying difficulty digital design assignments using the DEEDS system. The student grades for each session of digital design course served as the output variables for this study. The input variables included average idle time, total number of activities, average number of keystrokes, average time, and total related activity for every exercise during specific sessions. Using data from previous session, trained ML algorithms, and tested algorithms using data from next session.

Chung, Jae Young, and Sunbok Lee. [45], had employed ML using random forests to forecast dropout risk among students. 165,715 high school students' samples from 2014 National Education Information System (NEIS), a nationwide system for educational administration information connected via the Internet to about 12,000 fundamental as well as secondary schools, 17 city/provincial offices of education, and the Ministry of Education in Korea, served as source of data for this study. In terms of several performance criteria for binary classification, expective approach achieved specially well in forecasting dropout rate of students. The outcomes of investigation show the value of integrating big data from students' education with machine learning.

R. Ghorbani and R. Ghousi. [46], had utilized several types of ML classifiers, such as Random Forest, LR, ANN, KNN, SVM, XG-boost, DT, and NB able to check outcome of resampling approaches improved in solving imbalanced issue. Additionally, as model validation procedures, Shuffle 5-fold cross-validation and Random hold-out approaches were employed. It was evident from the obtained findings with various assessment metrics that models with fewer classes and nominal features will perform better. Additionally, classifiers struggle with unbalanced data, thus resolving this issue was essential. With balanced datasets, classifier performance was enhanced.

H. Pallathadka, *et al.* [47], had estimated achievement of student in a course depend on their past performance in similar courses. A group of methods known as "data mining" were employed to find patterns buried within enormous volumes of already-existing data. These trends could be provided to forecasting and analysis. Education data mining is the term used to describe the collection of applications that use data mining in the field of schooling. These programs were centered on the study of teacher and student data. Applications of the analysis include classification and forecasting. SVM, C4.5, ID3, and Naive Bayes machine learning techniques were examined. The UCI mechanical performance information for students set was used in this experimental investigation. A few metrics, such as accuracy and mistake rate, were used to analyze algorithms.

M. Yağcı, [48], had analyzed the midterm test results as the source data to propose a new model based on ML processes to predict final exam scores of undergraduate students. To forecast the students' final exam marks, the ML algorithms nearest neighbor, Naïve Bayes, logistic regression, SVM, k-nearest neighbor, and random forests were evaluated and their results compared. The academic accomplishment grades of 1854 students who enrolled in a Turkish public university's Turkish Language-I course during autumn semester of 2019–2020 made up dataset. According on results, proposed approach had a 70–75% classification accuracy. The three categories of parameters were used to make predictions: departmental data, faculty data, and midterm exam grades. Such data-driven research was crucial in helping to make decisions and in developing a framework for learning analysis in higher education. To sum up, this study finds machine learning techniques and makes a significant contribution to the early detection of pupils who were at an elevated risk of underperforming. The research's Table 2.1 lists comparisons and prediction models for student performance.

Table 2.1 Comparison of the related work based student performance prediction models

Authors	Techniques	Advantages	Disadvantages
Amjad *et al.*,[30]	Machine Learning	• Analysis reveals correlation among poor academic performance and social media usage.	• Fails to comprehensively analyze negative impact of technology usage. • When it came to their performance, students who used social media only on the weekends outperformed students who used it during the week. Additionally, the effect of additional characteristics on students' performance was need to be quantified.
Nti, I.K., *et al.*, [31]	Machine Learning	• Error-based method improves predictive accuracy in assessing student performance.	• Limited generalizability due to small dataset of 100 participants. • From a single department, it represented a large number of participants. • It had suggested that the model performance was slightly affected by the small data set used in this study. Thus, larger datasets need to be utilized in order to

			enhance model performance.
Poudyal et al. [32]	2D-CNN	• Offers scalability and flexibility. • It can used to predict academic achieve of student on numerical 1D educational datasets • The EDM brings hidden information that helps to analysis and high-accuracy prediction of students' passing or failing grades from such raw data.	• Loss due to pooling
Ghassen Ben Brahim [33]	RF, SVM	• Training will be fast. • The RF classifier generated best classification accuracy performance of 97.4% for our model, according to the results. • It demonstrated that it was possible to forecast student performance through examining through their interaction logs.	• Consume memory for large data.
Asselman *et al.* [34]	Adaboost, XGBoost and random forest ensembled algorithm	• Offer versatility by allowing the integration of multiple models into a single framework. • It significantly improved the performance when comparing the performance prediction to the original PFA method. • Despite the impressive enhancement that all ensemble learning models can provide, it was important to recognize that application of these classifiers needs specific	• Lower accuracy, complexity

		environmental conditions for optimal performance.	
Kusumawardani *et al.* [35]	LA technique based on deep learning	• Great potential for use in Learning Analytics (LA) • The weighted loss function with weekly feature aggregation produced better outcomes. • It discovered that the positional encoding-free model performed better in every OULAD scenario.	• Lower accuracy, high computational complexity, overfitting, and class imbalance
Rivas *et al.* [36]	CNN	• Automatically learn the features. • To determine the benefits of virtual learning with certainty, it helpful if it quantify amount of time students spend relating with this technology. • The significance of interrelating with virtual classroom was demonstrated in this work.	• Risk in overfitting
Rodríguez-Hernández et al., [37]	ANN	• ANNs demonstrated significant predictive power. • Research indicates that a student's academic outcomes in higher education can be significantly predicted by their earlier academic accomplishment, social status, and high school characteristics. • This study concludes with some suggestions for using artificial neural networks and some issues to consider	• Reliance primarily on SABER 11 test outcomes limits the scope of predictors, neglecting potentially valuable information such as high school grade point average

		to analyzing academic achievement in higher education.	
Shahzadi et al., [38]	Machine Learning	• Identifies negative result of excessive social media on theoretical outcomes. • The findings show a substantial correlation among academic achievement and social media use among students. • The outcomes of study demonstrate that extensive use of social media had an impact on academic achievement.	• Method's accuracy in predicting performance solely based on social media not 100% reliable.
Alshamaila *et al.*, [39]	Intelligent Approach	• Illustrates beneficial impact of YouTube-based supplementary education on performance. • It employed an innovative approach when examining the social media impacts undergraduate student academic performance.	• Lack of clarity in real data regarding duration and nature of mobile and social media usage.
Waheed *et al.*, [40]	Deep ANN	• The objective of study was to assist higher education decision-makers create framework that was required for pedagogical assistance in order to support sustainable education. • This work also shows how effectively the deep learning model predicts student performance in advance, allowing the university to promptly intervene and implement remedial plans for student support and	• Fails in enable those involved in education in creating the necessary pedagogical policies and procedures.

		counseling into effect.	
Giannakas, Filippos, *et al.* [41]	DNN	• When Adagrad and Adadelta optimizers were used, prediction accuracy of approach was found to be 76.73% and 82.39%, while its learning performance was 80.76% and 86.57%, indicating that it was capable of accurately and sufficiently anticipating the performance of teams. • The suggested structure was selected after an initial examination of various architectures of model.	• It need to investigate its capacity to forecast group formation support and to offer teams focused cooperative activities to maximize their performance.
Tsiakmaki, Maria, *et al.* [42]	DNN	• The results demonstrated that most of the time a reasonable performance was obtained, and the suggested strategy significantly outperforms the baseline model. • Using the dataset from an earlier course (the "source course"), transfer learning allows to train deep network and then utilize that same dataset to start a new, similar study.	• Investigating effectiveness of transfer learning in unbalanced datasets gathered from various educational contexts was needed.
Gray, Cameron C., and Dave Perkins. [43]	ML	• The existing student information systems at Bangor University had been improved with the use of this proposed approach. • It had selected a classifier and measurement combination through a	• A longterm research was required, both within Bangor and when comparing similar groups at other universities.

		series of trials that best achieves the early identification objective.	
Hussain, Mushtaq, *et al.* [44]	ANNs, SVMs	• Teachers should be able to report better student performance in next session because ANNs and SVMs were simple to incorporate into the TEL system. • Instead of waiting to study until the final exam, this work was completed more effectively throughout the class. • Teaching, learning, and student achievement can all be enhanced by utilizing an ANN or SVM.	• To assist teachers in improving the performance of students who were not as prepared, it was necessary to employ K-means approach to examine learning performances of students throughout their interfaces with DEEDS.
Chung, Jae Young, and Sunbok Lee. [45]	Random forest	• In this study, random forests approach performed exceptionally well in terms of forecasting dropout rates among students using different performance measures for binary classification. • The outcome of research determine how well schools can record, examine, and produce background information or observation of data to identify those who at risk of dropping out.	• The few descriptive features in this research were one of its shortcomings. • The NEIS data fails to include many of the factors that were covered during the evaluation.
R. Ghorbani and R. Ghousi. [46]	ML classifiers, SVM-SMOTE	• The efficiency of the SVM-SMOTE exceeds the performance of the other resampling techniques. • When random hold-out technique was applied to various balanced data produced by different	• To obtain greater performance and a better comparison, new ensemble and hybrid classifiers were missed in this approach. • Need to offer a more comprehensive understanding of

		resampling techniques, it can be shown that some classifiers perform better and that all of the classes were predicted, indicating that the classifiers' performance was sufficient.	significant features.
H. Pallathadka, *et al.* [47]	SVM, ID3, C4.5, and Nave Bayes	• An institution can raise the standard of its teaching with the help of this type of analysis. • Parameters including accuracy and error rate were used to evaluate algorithms. • The best method for categorizing a set of student performance data was support vector machines (SVM).	• Teachers were need to concentrate more on the students who require it when they use this type of analysis.
M. Yağcı, [48]	ML	• These types of data-driven research were essential to creating an approaches for learning analysis in advanced education and assisting in decision-making. • This study recognizes maximum effective ML methods and makes a contribution to early prediction. • It can be used to forecast students' final test marks because of its high accuracy rate.	• Only the midterm grade, instructor, and department parameters were used to make the forecast. • It had been carried by including additional machine learning algorithms into the modeling process and by adding more parameters as input variables.

2.3 RESEARCH GAPS

- The ML methodology is used in student performance analysis methods developed. Automatic feature extraction is not appropriate in this scenario, and feature extraction is achieved by humans before the classification assignment is completed.

- When it comes to the data used for analysis, machine learning techniques perform worse for larger datasets but produce better results when dealing with smaller amounts of data.

- A major factor in determining the method's accuracy is selection of important features. The accuracy of detection is impacted and incorrect outcomes are produced when irrelevant factors are taken into account when assessing the achievement of a student.

- The accuracy of student performance analysis is limited by various factors, including processing complexity, laziness in learning, failure in managing noisy data, and unpredictability in student performance classification.

- Predicting students' performance on online platforms is growing more difficult due to factors like lower learning ability, overfitting issues, and complicated inputs.

- The performance has been measured by statistical techniques using each student's historical data; however, the process of detecting patterns using statistical methods is difficult, error-prone, and time-consuming.

The researchers intend to use artificial intelligence methods to address these issues. To identify student academic outcomes with more accuracy and speed, a hybrid deep learning method is used.

2.4 SUMMARY

This section makes the entire research because it covers the different categorization strategies that are utilized for the different characteristics. Researchers can identify the weaker pupils in their educational institutions with the use of this review, and they can then take proactive steps to enhance their learning or study habits in order to perform better in the future. Therefore, it want to conduct some experimental work in the future utilizing student data in an effort to generate useful outcomes using these algorithms.

CHAPTER 3

DEEP AUTO ENCODER BASED ON A TRANSIENT SEARCH CAPSULE NETWORK FOR STUDENT PERFORMANCE PREDICTION

This chapter presents Deep Auto Encoder (DE) depend on a transient search capsule network for prediction of student performance. It includes about Data Pre-processing, feature extraction, Feature selection utilizing Fuzzy Equilibrium Optimizer (FEO) and feature classification. Also provided with results and discussions of the proposed model for student performance.

3.1 OVERVIEW

The education sector today is akin to a competitive business environment, where institutions vie for excellence and profitability. Amidst this, the paramount goal remains to provide excellence schooling for the holistic advancement of students and the nation. However, modern universities encounter challenges in ensuring student efficiency [49]. Identifying students' strengths and weaknesses is crucial for tailored support, leading to enhanced academic performance. To address these challenges, predictive analytics, particularly machine learning (ML) techniques, have gained traction. By analyzing various factors such as students' learning environment, teacher-student relationships, family background, and mental well-being, predictive models aim to forecast student performance early on. These models often rely on attributes like CGPA, internal assessments, previous grades, as well as demographic and socio-economic data [50].

Commonly used ML procedures like KNN and SVM have shown promise in performance prediction [51]. However, challenges persist, including attribute identification, prediction accuracy, and scalability. In response, DL strategies, including CNN, LSTM, and DNN, have emerged [52]. DL models exhibit automatic learning capabilities and excel in analyzing vast datasets, offering improved prediction accuracy. Despite their advantages, DL methods face challenges such as time complexity and performance degradation. To mitigate these issues, innovative approaches like capsule network-based deep autoencoders are being explored to enhance prediction accuracy while reducing complexity.

Hence, the integration of predictive analytics, ML, and DL techniques holds immense potential for revolutionizing student performance prediction in academia. By leveraging these technologies, educational institutions can proactively identify and support students, thereby fostering a conducive learning environment and contributing to individual and institutional growth.

3.2 PROPOSED METHODOLOGY

There are numerous barriers that hinder success, such as fear. Hence, the analysis of the learning procedure can be comprehended by categorizing attention levels of student as high, medium, as well as low utilizing various ML approaches. In this chapter, introduced the Hybrid TSCNDE to identify student execution.

- To begin with, collected and pre-processed the educational dataset of students in order to eliminate any replicate records.
- The extraction of feature process employed a DNN to examine the distinct aspects for performance of student. To simplify the learning process and avoid the complexity of numerous features, utilized a FEO for attribute selection.
- The TSCNDE method incorporates fuzzy rules and generates a feature matrix through a fuzzy system. For predicting performance of student academic, the fuzzy output is fed into the TSCNDE. This algorithm effectively reduces the number of features while preserving their original properties. Additionally, it reduces overall computation time as well as memory space.
- Furthermore, applied a hybrid model to obtain an initial solution in this combination. The models were tested using test data, and the outcomes were evaluated dependent on parameters such as specificity, accuracy, precision, and sensitivity.

To calculate accuracy, utilized a confusion matrix as it is a reliable metric for examining unbalanced data. This method also helps in reducing time complexity.The TSCNDE method was implemented using Python.

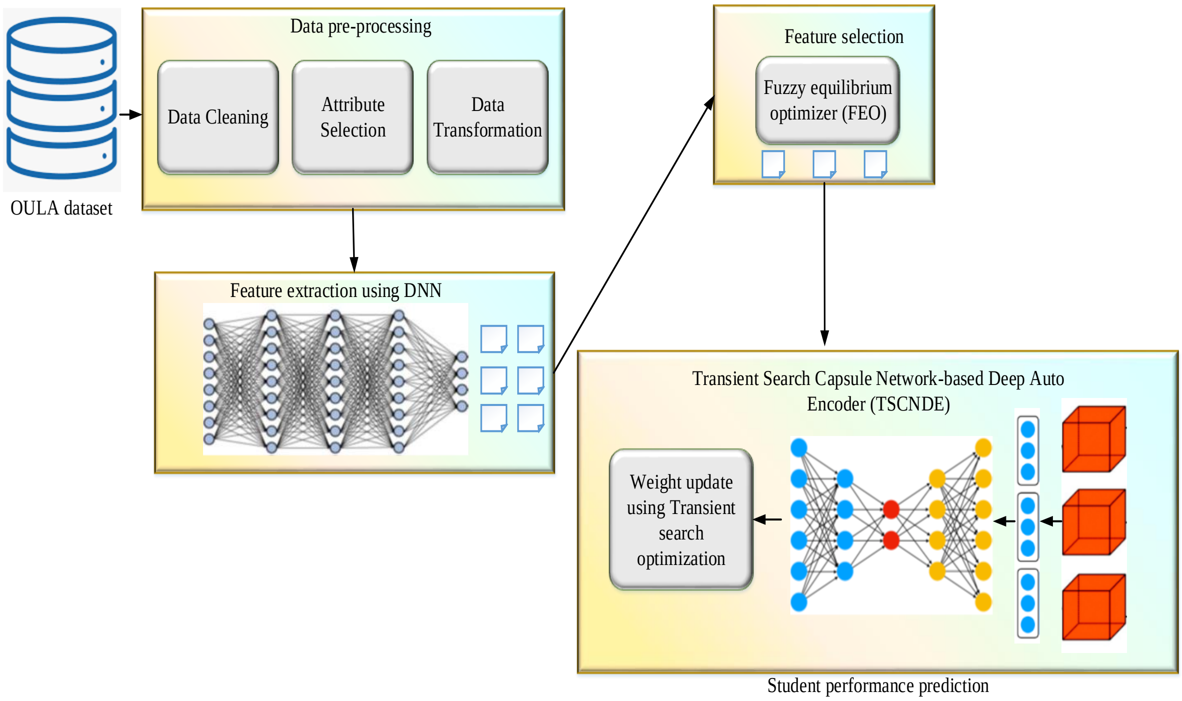


Fig. 3.1 Block diagram of the TSCNDE method

3.2.1 Data Pre-processing

The first and crucial procedure in forecasting student performance involves data pre-processing, a data mining technique aimed at refining raw data for further analysis. The OULA dataset represents a substantial volume of data, encompassing information about students in Learning Management Systems (LMS) along with interactive logs from virtual learning programs. However, this raw data often contains redundancies or unrelated information, necessitating data pre-processing to ensure its suitability for analysis.

Data pre-processing consists of several key phases, including data cleaning, attribute selection, and data transformation, each playing a vital part in enhancing the quality and usability of the dataset.

- Data Cleaning: This initial step involves the removal of irrelevant or noisy elements from the dataset. Noise, missing data, or inconsistencies within the dataset are identified and addressed to improve data quality. Techniques such as patching, reduplication, or removal of incomplete data entries are employed to ensure the dataset is accurate and complete.

- Attribute Selection: The presence of numerous feature attributes can diminish the efficiency of predictive models. Attribute selection goals is to identify and retain only the most pertinent and impactful features for performance modeling. By reducing the number of attributes, attribute selection helps streamline the analysis process and improve the model's predictive accuracy.

- Data Transformation: Once the data is cleaned and relevant attributes are selected, it undergoes transformation to prepare it for analysis. This stage involves translating the info into a suitable format for further processing. The transformed data may still contain a large volume of information, making it challenging to handle efficiently. Therefore, data reduction techniques are applied to condense the dataset into a more manageable form while preserving its essential characteristics. This final stage of data reduction ensures that the output is conducive to subsequent feature extraction processes.

Overall, data pre-processing is a critical precursor to predictive modeling in the context of student performance prediction. By refining raw data through cleaning, attribute selection, and transformation, obtained a high-quality dataset that facilitates accurate and insightful analysis, eventually leading to more effective predictions and interventions in educational settings.

3.2.2 Feature Extraction

Extraction of feature plays a crucial role in reducing the dimensionality of datasets containing redundant information, thereby facilitating faster data processing and reducing system complexity. In the context of the OULA dataset, extraction of feature is performed utilizing a DNN model, a powerful algorithm widely utilized in deep learning applications. The architecture of a DNN comprises an first layer, hidden layers, as well as an resultant layer. The input layer obtains the data, which is then approved through hidden layers where computations occur, leading to valuable feature extraction. Finally, the output layer offers the desired information depend on the processed data. In the case of the OULA dataset, which is employed for predicting student performance, eleven attributes are initially considered, but 2 attributes are deemed redundant. The resting attributes are mined utilizing the DNN model.

The extraction process involves configuring the input layer to receive attribute information and passing it through hidden layers for feature calculation. The hidden layers, depending on the complexity of the algorithm, may consist of multiple layers with multiple nodes each. Through computations performed within these layers, relevant features are filtered+ from the set of informations, contributing to the prediction of student performance.

$$
\begin{aligned}
O^1 &= \{O_1^1, O_2^1, ..., O_x^1, ...O_3^1\}; 1 \le x \le 3 \\
A^1 &= \{A_1^1, A_2^1, .., A_o^1, ..., A_e^1\}; 1 \le o \le e
\end{aligned}
\tag{3.1}
$$

where O_x^1 stands for the observable neuron that is the x^{th}. The hidden layers are denoted by A^1, the hidden neuron is indicated by o^{th}, and the total amount of hidden neurons is denoted by e. The weighted value as well as bias, which are expressed as follows, were calculated by the HL nodes during their design.

$$
P_j = \sum_i^m k_i * w_{ij} + \theta_j \tag{3.2}
$$

The input data in this case is k_i, the bias is θ, and the weighted link is w_{ij}. Consequently, the transfer function (sigmoid) that follows may be used to convert P_j:

$$
F(P_j) = \frac{1}{1 + e^{-P_j}} \tag{3.3}
$$

Subsequently, the network training process begins with initializing the weights by selecting a range of random values. The network weights are modified as necessary during training:

$$
\Delta w_{ij} = -n \frac{\theta E}{\partial w_{ij}} \tag{3.4}
$$

where the error is represented by E and the learning rate by n. In this case, M represents the state activation function. The following equations may be used to denote all DNN procedures. The operation for input and output is provided by,

$$
i_t = M(k_t\, a_t^i + h_{t-1}\, w_t + \theta_i) \tag{3.5}
$$

$$
o_t = M(k_0\, x_t^i + h_{t-1}\, w_0 + \theta_0) \tag{3.6}
$$

If the weight updated from the hidden vector is w, the input-output states are represented by i_t *and* o_t, and the bias vector is θ.

3.2.3 Feature selection using Fuzzy Equilibrium Optimizer (FEO)

FEO is highly useful in evading the complexity of a number of features in the learning procedure and is utilized for the attribute selection procedure. The fuzzy rule feature selection procedure may be explained as follows.
Let Y be the data that is used to choose the features; it may be expressed in the following way.

$$Y = \{ y_d \mid d = 1, 2, 3, ..., n \} \tag{3.7}$$

In this case, n denotes the total number of patterns, $y_d \in R^a$ and a the amount of characteristics f_i. Let y_d^a be the value of pattern y_d's a^{th} feature, f_a. Then, by utilizing a vector, we may define every pattern in y_d the labeled dataset Y.

$$y_d = [y_d^1, y_d^2, y_d^3, ..., y_d^a] \tag{3.8}$$

Let F represent a set such that every member in F corresponds to a feature. That is,

$$F = \{ f_1, f_2, f_3, ..., f_a \} \tag{3.9}$$

In this case, project the dataset onto a fuzzy space using a membership function H. The following is a representation of the membership function H.

$$H = \{ \beta_{11}, \beta_{12}, \beta_{13}, ...\beta_{1p}, \beta_{21}, \beta_{22}, \beta_{23}, ..., \beta_{2q}, \beta_{a1}, \beta_{a2}, \beta_{a3}, ...\beta_{ar} \} \tag{3.10}$$

where the j^{th} fuzzy set of the i^{th} feature is characterized by β_{ij}. The cardinality of the fuzzy sets for the first, second, and q^{th} features is represented by the positive real numbers denoted by q, p, r in the membership definition set. Let F_x

represent the dataset Y's predicted collection of fuzzy values. It has the following definition. where $\beta(y_d)$ is a vector that equation (3.12) uses to represent it.

$$F_x = (y_d, \beta(y_d)) \tag{3.11}$$

$$\beta(yd) = [\beta_{11}(y_d^1), \beta_{12}(y_d^2), ..., \beta_{1p}(y_d^1), \beta_{21}(y_d^2), \beta_{22}(y_d^2), ...\beta_{a1}(y_d^a), \beta_{a2}(y_d^a), ...\beta_{ar}(y_d^a)] \tag{3.12}$$

Let $|f_i|$ represent how many fuzzy sets there are for every f_i characteristic. Thus, the fuzzy set β_{ij} has the following definition.

$$\beta_{ij} : y_d^i \rightarrow [0,1]; \forall i \in \{1,2,3,...a\} \wedge \forall j \in \{1,2,...|f_i|\} \wedge k \in \{1,2,3,........n\} \tag{3.13}$$

Using the value $\sum_{i=1}^{a} |f_i|$, the resolution of membership function is quantified. The fuzzy selection procedure is determined by the best possible mixture of a fuzzy set (β_{ij}). Let α represent every conceivable combination of a fuzzy subset and $j(.)$ be the function utilized to assess a subset of a fuzzy set. The next step is to produce a fuzzy set subset $H_{optimal}$ that must meet the next equation in order to do fuzzy feature selection.

$$j(H_{optimal}) = E(j(H_i)), \forall H_i \subseteq \alpha, \alpha = 2^H \tag{3.14}$$

In this case, E represents the minimum or maximum operator. An Equilibrium Optimizer (EO) can be used to adjust the value of $H_{optimal}$. The following is a description of the necessary optimization procedure for the feature selection.

3.2.3.1 Parameter optimization using EO

The four steps listed below outline the mathematical modeling of EO.
Step 1: *Initialization*

EO makes use of a set of components, every of which represents a vector of awareness. The unplanned answer for $H_{optimal}$ is included in this concentration vector. The following equation describes the initial concentration vector's random creation.

$$v_{ci} = c_{\min} + (c_{\max} - c_{\min}) * s_i \quad i = 1,2,...n \tag{3.15}$$

Whereas v_{ci} represents the concentration vector, $c_{\min}$ determines the problem's upper bound dimension, and $c_{\max}$ determines the problem's lower bound dimension. Equation 15 shows that s_i is a random vector with a [0,1] interval, and n is the overall amount of particles that make up the population inside the group. s_i has a value that ranges from 0 to 1.

Step 2: *Equilibrium pool and candidates (* E_{pc} *)*

Like any optimization algorithms, EO also aims to produce a superior optimization outcome. It constantly looks for the equilibrium state of the system. Once equilibrium is reached, it is compelled to proceed in the direction of the almost optimum solution to the optimization issue. The level of concentration needed to reach equilibrium throughout the optimization procedure is unknown to EO. It is hence compelled to allocate five particles. Four of the 5 units are the top in the populace; the last particle is the normal of these four. These five equilibrium particles are utilized in the operator's further exploitation and investigation. The five chosen particles are kept in vector storage, sometimes referred to as equilibrium pools. The equilibrium pool is often represented by the equation that follows. The stages comprised in parameter optimization with the EO technique are outlined in Algorithm 1.

$$\vec{E}_{pc} = [\vec{E}_{pc(1)}, \vec{E}_{pc(2)}, \vec{E}_{pc(3)}, \vec{E}_{pc(4)}, \vec{E}_{pc(avg)}] \tag{3.16}$$

Algorithm 1: Parameter optimization using EO

```
Initialize the population of a particle (i= 1, 2, 3……….. n)
Set the fitness value of four particles in the equilibrium pool, O_pc
with a large value.
Set the values of parameters
While (n_it < τ_max)
    For each i^th particle
    Calculate the fitness value of the particle i A_f (O_pc(i))
```

If $\left((\vec{A}_f(\vec{O}_{pc(i)}) < (\vec{A}_f(\vec{O}_{pc(1)})\right)$

Set $\vec{O}_{pc(1)}$ with $\vec{O}_{pc(i)}$ and $\vec{A}_f\,(\vec{O}_{pc(1)}\ \mathit{with}\ \vec{O}_{pc(i)})$

else If $\left(\vec{A}_f(\vec{O}_{pc}) > \vec{A}_f(\vec{O}_{pc(1)}\ \mathit{and}\ (\vec{A}_f(\vec{O}_{pc(i)}) < \vec{A}_f(\vec{O}_{pc(2)})\right)$

Set $\vec{O}_{pc(2)}$ with $\vec{O}_{pc(i)}$ and $\vec{A}_f\,(\vec{O}_{pc(2)}\ \mathit{with}\ \vec{O}_{pc(i)})$

else If $\left(\vec{A}_f(\vec{O}_{pc}) > \vec{A}_f(\vec{O}_{pc(2)}\ \mathit{and}\ (\vec{A}_f(\vec{O}_{pc(i)}) < \vec{A}_f(\vec{O}_{pc(3)})\right)$

Set $\vec{O}_{pc(3)}$ with $\vec{O}_{pc(i)}$ and $\vec{A}_f\,(\vec{O}_{pc(3)}\ \mathit{with}\ \vec{O}_{pc(i)})$

else If $\left(\vec{A}_f(\vec{O}_{pc}) > \vec{A}_f(\vec{O}_{pc(3)}\ \mathit{and}\ (\vec{A}_f(\vec{O}_{pc(i)}) < \vec{A}_f(\vec{O}_{pc(4)})\right)$

Set $\vec{O}_{pc(4)}$ with $\vec{O}_{pc(i)}$ and $\vec{B}_f\,(\vec{O}_{pc(4)}\ \mathit{with}\ \vec{O}_{pc(i)})$

end if

end for

$\vec{O}_{pc(avg)} = (\vec{O}_{pc(1)} + \vec{O}_{pc(2)}, +\vec{O}_{pc(3)} + \vec{O}_{pc(4)})/4$

The equilibrium pool $\vec{O}_{pc} = [\vec{O}_{pc(1)}, \vec{O}_{pc(2)}, \vec{O}_{pc(3)}, \vec{O}_{pc(4)}, \vec{O}_{pc(avg)}]$

Accomplish the memory saving

Assign the value of τ using $\tau = \left(1 - \frac{n_{it}}{\tau_{\max}}\right)^{\left(d_o * \frac{n_{it}}{\tau_{\max}}\right)}$

For each i particle

Choose one candidate from the $\vec{O}_{pc}$

Generate random vector γ

Construct $\vec{A}_f$ using $\vec{A}_f = e^{-\vec{\gamma}(\tau-\tau_0)}$

Update the concentration

end for

n_{it}^{++}

end while

Step 3: *Updating the concentration*

EO often strikes a balance among intensification and diversity. Let $\vec{\gamma}$ be a casual vector that falls inside the range [0, 1]. The fitness function expression then makes sense as follows.

$$\vec{A}_f = e^{-\vec{\gamma}(\tau-\tau_0)} \tag{3.17}$$

In this case, τ denotes the current iteration, while τ_0 represents the starting value. An exponential term is signified by $\vec{B}_f$. The subsequent equation shows that the value of τ declines as the number of iterations rises.

$$\tau = \left(1 - \frac{n_{it}}{\tau_{\max}}\right)^{\left(d_o * \frac{n_{it}}{\tau_{\max}}\right)} \tag{3.18}$$

In this case, n_{it} stands for the overall amount of iterations, τ and $\tau_{\max}$ for the maximum and current values of iterations, respectively. Constant parameter d_0 is utilized to adjust intensification capability.

Step 4: *Parameter optimization*

By updating the concentration, the parameter may be optimized. The process of parameter optimization is represented by the following equation.

$$H_{opt} = \vec{O}_{pc} + (\vec{O} - \vec{O}_{pc}) * \vec{A}_f + (\frac{1}{\gamma * V}) * (1 - \vec{A}_f) \tag{3.19}$$

Suppose that γ is a random vector ranging from [0, 1]. F is an exponential term, and V=1.

3.2.4 Feature classification

The procedure of classifying a vast quantity of data into distinct groups in order to identify patterns is known as feature classification. The procedure of feature selection is always followed by feature categorization. In the feature classification step, the chosen characteristics are grouped based on a few criteria. There are several approaches to going through the categorization process. The process of categorization guarantees more precise outcomes. Here, the TSCNDE is utilized to aid with the feature categorization process. The TSCNDE receives the fuzzy output in order to forecast the academic success of the students.

3.2.4.1 Capsule network Deep Autoencoder

In general, a capsule represents a vector whose length reflects the likelihood that an entity exists. By identifying every aspect that is relevant to students, capsules are grouped like a collection of neurons to aid in the analysis of the entire feature set. The entity's attributes are indicated by the vector's orientation. The usual design of a deep autoencoder based on a capsule network is seen in Fig. 3.2.

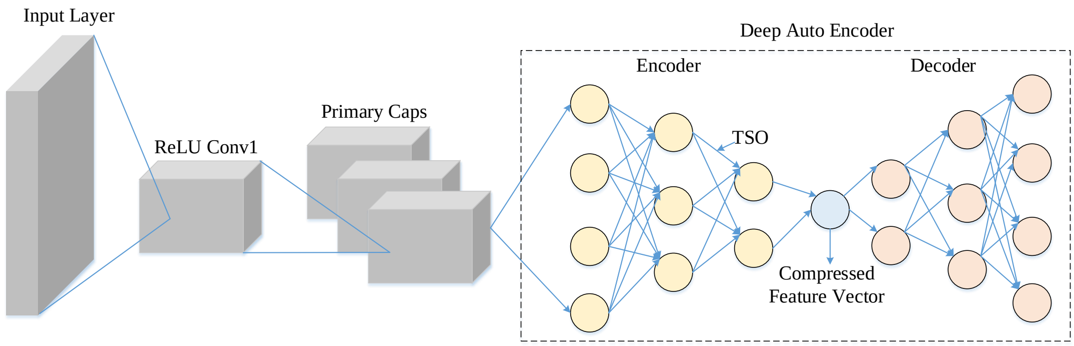


Fig. 3.2 Architecture of Capsule Network-based Deep Auto Encoder

A capsule network is composed of four major components, as shown in figure 3. Convolution, Primary caps, Digit caps, as well as decoder are these layers. Here, the Deep Autoencoder (DAE) is used to replace the last component, which is the Digit Caps and decoder. Let i_j represent the input characteristics that were applied to the capsule network, and let o_j represent the matching output. The result operation of the capsule network is represented by the following equation.

$$o_j = \frac{\|i_j\|^2}{1+\|i_j\|^2\|i_j\|} \tag{3.20}$$

Initialization parameters (IP) are a collection of vector values pertaining to student attributes that make up the capsule network's output. Owing to certain constraints, these parameters do not yield a precise outcome. As a result, the Transient Search Optimization (TSO) technique, another optimization procedure, is used in this case to maximize the capsule network's output parameters.

3.2.4.2 Parameter Optimization using TSO algorithm

TSO is a physics-based meta-heuristic algorithm [53]. Here, TSO is applied to improve the capsule network's output parameter based on the transient performance of switching circuits. The following features of this TSO method, like search-agent initialization amongst search area lower as well as upper bounds, optimal searching (exploration stage), and locating the best (exploitation stage) solution, may be used to model this approach. The following is the algorithm used to carry out the TSO procedure. Algorithm 2 mentions the TSO parameter optimization.

Algorithm 2: Parameter optimization using TSO

Initialize the population and the finest locations
Estimate the maximum number of iterations in the population
While $l < l_{\max}$
    Update the values of O_j utilizing equation (3.22)
do for all populations
    Update the population
end do
    Compute the number of iterations of all new populace
    Inform the finest rate if the iteration value is less than the earlier value.
$l = l + 1$
end while
Outcome the optimized value

The TSO search agents are first produced and initialized at random as,

$$Z = l_b + ran \times (u_b - l_b) \tag{3.21}$$

Next, the oscillation of the 2nd imperative RLC circuits around zero drives the TSO exploration performance, while the exponential decay of the first order discharge characterizes the TSO exploitation. The TSO algorithm's accurate model containing both exploration and exploitation may be written as follows:

$$Z_{l+1} = \begin{cases} Z_l^* + (Z_l - Z_l^* \cdot C_1)e^{-T} & r_1 < 0.5 \\ Z_l^* + e^{-T}\left[\sin(2\pi T) + \cos(2\pi T)\right]\left|Z_l - Z_l^* \cdot C_1\right| & r_1 \geq 0.5 \end{cases} \tag{3.22}$$

$$T = 2 \times v \times r_2 - v \tag{3.23}$$

$$C_1 = v \times k \times r_3 + 1 \tag{3.24}$$

$$v = 2 - 2\left(l \Big/ L_{\max} \right) \tag{3.25}$$

Where ran is the random value, l_b and u_b are the lower as well as upper bounds, C_1 and T are the random coefficients, v is the variable that fluctuates between 2 and 0, and r_1, r_2, r_3 is the random number. The search agent position is Y_l, the maximum number of iterations is $L_{\max}$, the iteration number is l, the best position is Y_l^*, and the constant number is represented by k as (k=0,1,…). The following equation (22) in TSO is taken into consideration in order to optimize o_j. The input of the DE receives this output. A feed-forward neural network is represented by DE, which produces duplicates of the input as the output. DE is a system with completely linked network layers, as seen in the image. There are two primary levels in DAE. At the bottleneck layer, the encoder section converts the high-dimensional data into a lower-dimensional representation. At the decoder, the bottleneck features are converted to produce high dimensional features. Equation (3.26) illustrates how an autoencoder (AE) might grasp the identity function.

$$D' = h_{w,b}(D) \approx D \tag{3.26}$$

In this case, D stands for the input, b for the bias, and w for the weight. The input D is then compressed into a form by the AE.

$$S = \eta(wD + b);\ \eta(x) = \frac{1}{1 + \exp(-x)} \tag{3.27}$$

where S represents the compressed form, which is written as the mapping function and decoded to provide output D'.

$$D' = \eta(Sw' + b') \tag{3.28}$$

The procedure can be repeated until $D \approx D'$, at which point the output of the final encoder represents the compact, high-level version of the input data. Here, the original characteristic is maintained but the maximal features are decreased.

3.3 RESULTS AND DISCUSSION

The subsequent is a picture of the output of the TSCNDE. The OULA dataset is used to assess student performance [54]. It includes both personal information and statistics on the past performance of the pupils. The implementation tool of choice is Python. Table 3.1 provides a description of the experimental conditions. The OULA dataset is briefly described in Section 3.1.

Table 3.1 Experimental configurations

Parameter	**Value**
EO_Generation Probability (GP)	0.5
Processor	Intel (R) Core (TM) i3-3220 CPU @ 3.30 GHz
RAM	8GB
Random numbers	$r_1 = r_2 = [0,1]$
System type	64 bit Operating System
TSO	a decreases from 2 to 0, k=0

3.3.1 OULA dataset

The Open University Learning Analytics set of information, or OULA for short, was used for the present research (https://analyse.kmi.open.ac.uk/open_dataset). This dataset includes many student-related characteristics. Among the features are clickstream history, student demographics, assessment submission details, and more. It is a raw tabular data set including information on 32,593 pupils over the period of nine months. The dataset was gathered between the years 2014 and 2015. Information on how the learner interrelates with the VLE environment is also included. The OULA dataset's details are shown in Table 3.2.

Table 3.2 Dataset Details

	StudentInfo	**StudentAssessment**	**StudentVIe**	**StudentRegistration**
Rows	32,593	173,912	10,655,280	32,593
	Age_band	score	date	date_unregistration
	code_module	id_student	code_presentation	code_presentation
	code_presentation	id_assessment	code_module	code_module
	disability	is_banked	id_site	date_registration
	gender		sum_click	
	highest_education	date_submitted	id_student	id_student

Columns	id_student				
	imd_band				
	Num_of_prev_attempts				
	region				
	Studied_credits				
	Final_result				

3.3.2 Experimental results

In this case, assessment metrics including sensitivity, specificity, accuracy, and precision are utilized to evaluate the TSCNDE technique. The four metrics examined here are seen to be highly useful for assessing various approaches. Table 3.2 presents the assessment outcome of TSCNDE in comparison to various current approaches, including KNN, DT, NB, and ES. This model allows for the achievement of the original solution. Equations (3.29), (3.30), (3.31), and (3.32) that follow show how to calculate several metrics, comprising sensitivity, accuracy, specificity, as well as precision.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.29}$$

$$sensitivity = \frac{TP}{TP + FN} \tag{3.30}$$

$$specificity = \frac{TN}{FP + TN} \tag{3.31}$$

$$precision = \frac{TP}{TP + FP} \tag{3.32}$$

The variables TN as well as TP represent the number of genuine negative as well as positive performance predictions in these equations (3.29, 3.30, 3.31, and 3.32). The false negative as well as false positive performance forecasts are denoted by the letters FN and FP. The amount of false positives generated from negative examples in the segmentation is shown by FPTN.

Table 3.3 Performance comparison of TSCNDE method with existing ML methods

Parameter	**TSCNDE (proposed method)**	**KNN**	**NB**	**DT**	**ES**
Accuracy (%)	**99.2**	98.2	98.2	97.8	98.7

Precision (%)	**99.8**	70	64	59	79
Sensitivity (%)	**98.9**	98.1	98.4	97.9	98.7
Specificity (%)	**98.7**	92.1	86.6	89.9	94.8

The comparison graph for several machine learning approaches is shown in Table 3.3. Here, the comparison procedure is carried out using the following four machine learning techniques: KNN, DT, NB, and ES. KNN is a regression and classification method. Similarly, the Decision Tree classifier (DT) and the Ensemble Stacking classifier (ES) are widely employed in the majority of classification procedures. The Naïve Bayes classifier, or NB, is said to be probabilistic and contributes to a better degree of accuracy. It is thought that all four of the strategies utilized in this comparison procedure are highly effective in achieving higher accuracy. Nevertheless, they don't produce significantly superior outcomes because of certain limits in other aspects. In Fig. 3.3, the TSNDE method's performance is shown visually. The vertical axis signifies the outcome of the TSCNDE approach, while the horizontal axis represents the many evaluation criteria taken into account for performance prediction.

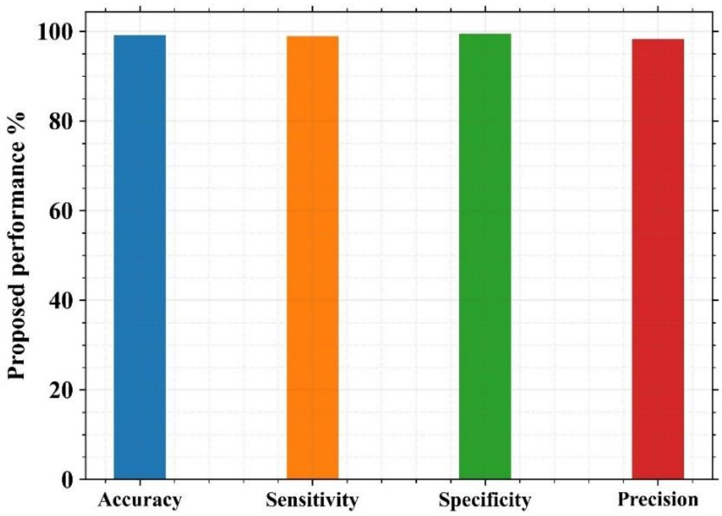


Fig. 3.3 Performance outcomes of TSCNDE method in parameters like Accuracy, Sensitivity, Specificity, and Precision

The TSCNDE method's performance is shown in Fig. 3.3. Numerous factors were examined, including precision, sensitivity, specificity, and accuracy. The TSCNDE technique has an accuracy of 99.2% when calculated, which is greater than any other approach currently in use. It suggests that the value is more in line with the pupils' assessed performance. The TSCNDE technique has a 98.9%

sensitivity. The TSCNDE approach has a somewhat higher sensitivity than the other methods that are currently in use, including NB and ES. It performs better than the DT approach. The TSCNDE technique's specificity is 98.7% after examination. Furthermore, a confusion matrix is also employed in this instance to evaluate performance.

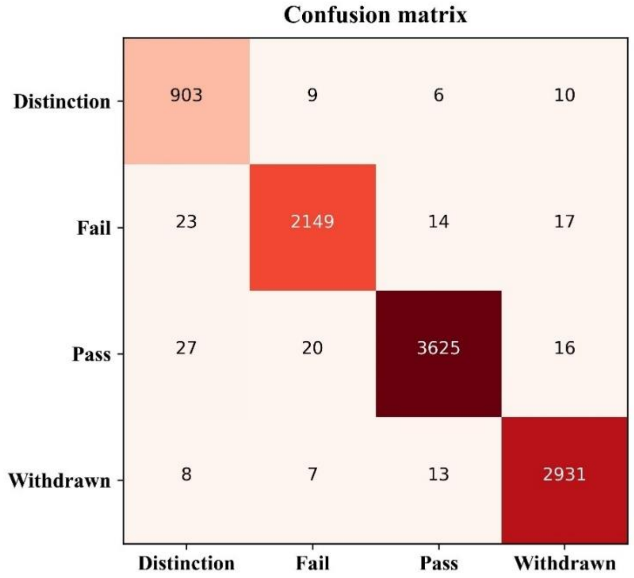
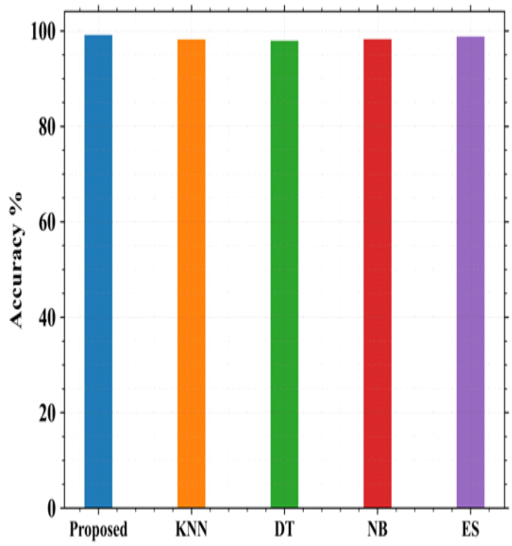
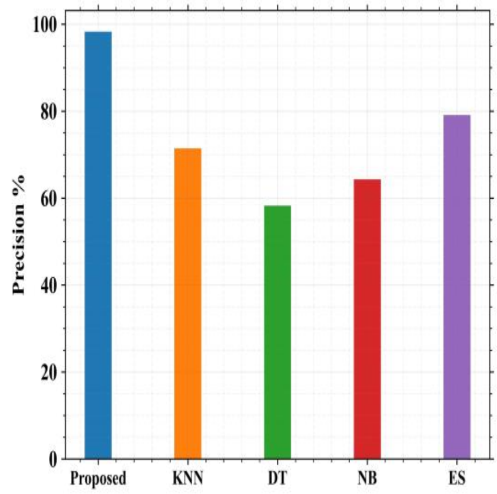


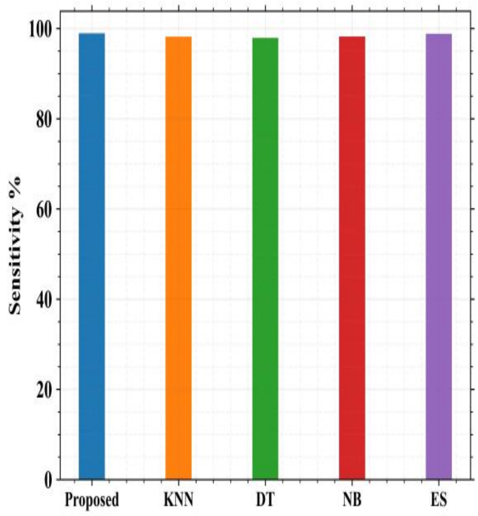
Fig. 3.4 Performance outcome using a confusion matrix

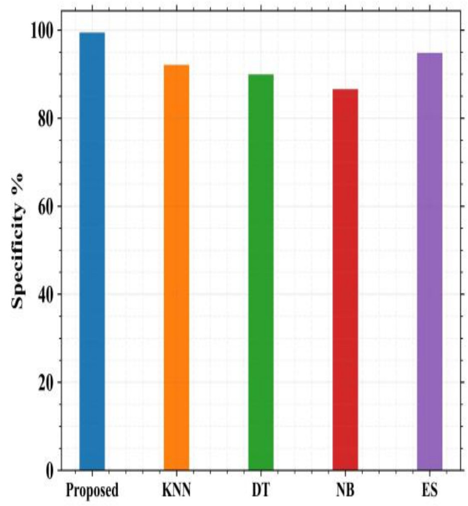
Fig. 3.4 shows the performance outcome utilizing the confusion matrix. The confusion matrix is a unique kind of arrangement with two dimensions: present and anticipated events. Such a table arrangement can aid in the conceptualization of the deep learning performance technique. Another name for the confusion matrix is the mistake matrix. In this matrix, each component in the column represents the real incidences, and each element in the row represents the anticipated occurrences (or vice versa). Fig. 3.4's confusion matrix provides an overview of the terms distinction, fail, pass, and withdrew. The link between the actual and anticipated class parameters is demonstrated. There are 903, 2149, 3625, and 2931 students who passed, withdrew, and received distinctions. Similarly, the confusion matrix may be used to accurately predict student execution. The contrast of the TSCNDE approach with current machine learning techniques is mentioned in Fig. 3.5.





(a) (b)





(c) (d)

Fig. 3.5 Performance comparison with ML techniques

The TSCNDE model is contrasted with various KNN, DT, NB, and ES techniques that are currently in use in Fig. 3.5. Four assessment metrics are used to take into account the comparison results: sensitivity, specificity, accuracy, and precision. The compassion of the TSCNDE method is better, as the figure mentioned. Compared to the DT approach, its performance is superior. In terms of sensitivity, it performs somewhat better than the KNN and NB approaches. The TSCNDE technique has 99.8% precision. In comparison to the DT, NB, and KNN methods, it performs better. The accuracy values of these three techniques are all less than 70%.

When it comes to predicting student achievement, the ES approach is over 80% accurate. It can be concluded by examining all of the precision-related data that the TSCNDE approach performs better in terms of precision.

For performance outcomes, the TSCNDE provides 99.2% accuracy. All other approaches are not as precise as our TSCNDE technique, as the figure illustrates. The TSCNDE technique has 98.7% specificity. Compared to the KNN, DT, and NB approaches (92.1%, 89.9%, and 86.6%), this specificity number is significantly superior. The specificity value of the TSCNDE technique is somewhat higher than that of the ES (94.8%) approach. In summary, compared to other approaches already in use, the TSCNDE technique achieves better in terms of accuracy, sensitivity, precision, and specificity.

3.3.3 Comparison with Deep Learning methods

The performance comparison of the TSCNDE approach with a few of the current DL techniques is shown in **Table 2.1**Table 3.4. To relate the TSCNDE technique's execution, three approaches are selected. LSTM is a DL technique applied to many classification procedures. Although this approach produces superior outcomes, such systems are thought to be exceedingly complicated and have large operational costs. Recurrent neural networks, or RNNs, are frequently utilized for various prediction procedures, as are DBNs (Deep Belief Networks). In the same way, autoencoders are a type of neural network structure that are commonly employed in numerous prediction tasks.

Table 3.4 Performance comparison with DL methods

Method	Accuracy (%)	Precision (%)
TSCNDE (proposed method)	**99.2**	**99.8**
LSTM	96.48	96.14
DBN	94.69	93.26
RNN	93.46	93.9
Autoencoder	92.89	91.79

The outcomes of the TSCNDE technique are displayed in Table 3.4 together with the current DL techniques, including LSTM, RNN, as well as Autoencoder. The reasonable examination on the OULA dataset is shown in Table 3.5.

Table 3.5 Comparison on OULA dataset

Author Name	Technique Used	Performance (%)
Hassan et al. [55]	LSTM	Accuracy (97.25%)
Heuer and Breiter [56]	SVM	Accuracy (90.85%)
Hlosta et al. [57]	ML	F1-score (71.31%)
Hussain et al. [58]	ML	Accuracy (88.52%)
Rizvi et al. [59]	DT	Accuracy (83.14%)
Proposed	**TSCNDE**	**Accuracy (99.2%)**

Numerous techniques for predicting student success are indicated in the graph. Due to their limitations, all three approaches perform poorly overall but produce superior outcomes in particular areas. The comparison of TSCNDE with a few of the deep learning techniques currently in use is shown in Fig. 3.6.

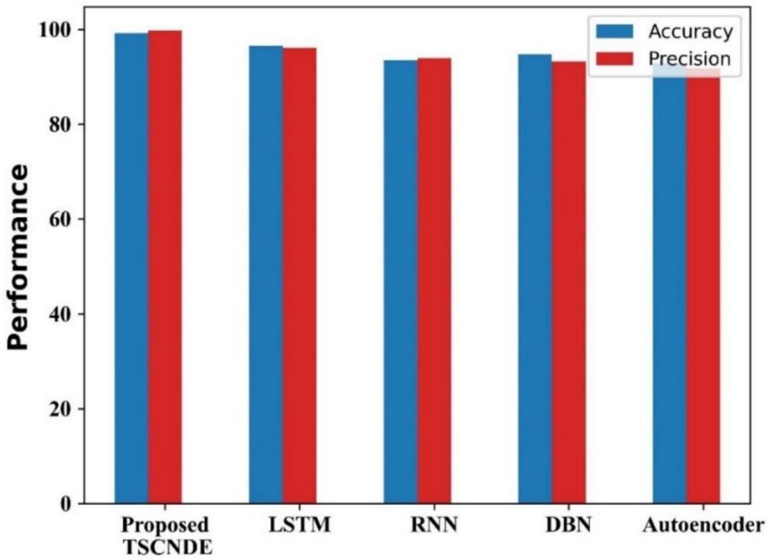


Fig. 3.6 Performance comparison with DL methods

The accuracy and precision comparison graph between the TSCNDE method and other DL techniques, including LSTM, RNN, and auto-encoder, is displayed in Fig. 3.6. The graph indicates that the TSCNDE approach has a 99.2% accuracy rate. It performs significantly better than other deep learning techniques as autoencoder (92.89%), LSTM (96.48%), RNN (93.46%), and DBN (94.69%). The TSCNDE method's accuracy value is 99.8%. LSTM, RNN, DBN, and Autoencoder have equivalent values of 96.14%, 93.9%, 93.26%, and 91.79%, in that order. In other words, TSCNDE outperforms other DL techniques. A ROC curve is shown in Fig. 3.7 to illustrate the outcome of the TSCNDE approach.

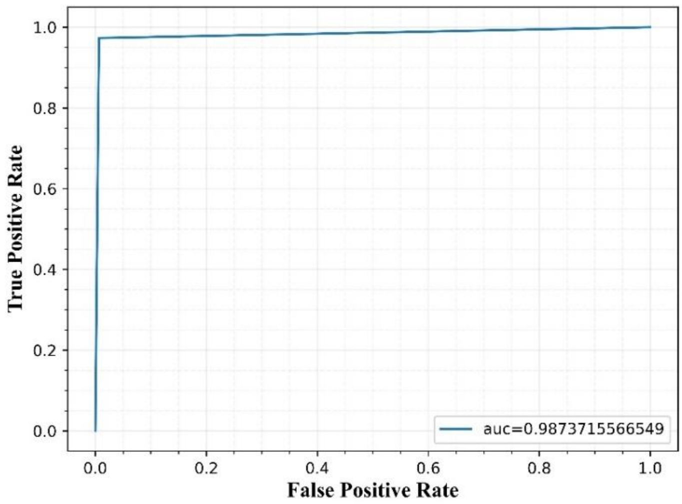


Fig. 3.7 ROC curve

The ROC curve of the TSCNDE technique is shown in Fig. 3.7. It is employed in this instance to assess the effectiveness of the TSCNDE categorization strategy. In the ROC curve, the True Positive Rate is mentioned by the Y-axis, while the X-axis represents the False Positive Rate. Different threshold values are used to aid with the performance evaluation. It demonstrates how the categorization process may distinguish between distinct classes. The execution of the technique grows as the curve's value upsurges.

3.3.4 Discussion

The advent of technology has made educational institutions unique in today's world, where they are essential to the growth of the nation. Education alters people's lives, as well as communities and the whole planet. The traditional educational system today includes additional components including online and web-based learning, workshops, seminars, and more. The majority of current approaches fail when dealing with the large quantities of data kept in educational databases. The execution forecast of learners with the achievement attained from student learning (SL) is the main objective of the suggested TSCNDE based DL framework. The academic performance of the learner is one of the most significant aspects of education. It is appropriate to detect low-performing kids by effective monitoring of their academic performance. The approach that helps the low-performing pupils in the beginning is the most appropriate: student performance prediction. The focus of mining educational data is to use deep learning models to discover interesting patterns from the OULA dataset. This work uses a TSCNDE to provide an effective prediction of student execution. Students' performance can be classified as Distinction, Fail, Pass, or Withdrawn using the suggested (TSCNDE) methodology. Precise forecasting of student performance is essential for enhancing instructional effectiveness, encouraging customized education, addressing the issue of student attrition, and more. Moreover, the suggested TSCNDE model has resulted in the successful prediction of data because of the notable advancements in Deep Learning (DL) models. Failure risk can be successfully reduced by making early predictions about the student's performance. To assess the benefits of the suggested approach, the outcomes on the OULA information set are compared using ML and DL algorithms.

The four metrics that are utilized in computing are specificity, sensitivity, accuracy, and precision. The outcomes demonstrated the importance of the suggested DL model in terms of early academic success prediction. This suggested approach predicts student achievement and serves as the foundation for early intervention in the event that students may not meet the learning objectives. The TSCNDE model that is being given attempts to examine how ML and DL approaches are used to predict student accomplishment in a methodical manner. For everyone involved in choosing the best tactics to support students in completing the learning objective, the current methods for forecasting the effectiveness of SL are quite important. Although machine learning models are widely employed, they are unable to manage large amounts of data. The challenge of identifying pupils who do worse academically or educationally and determining the students' performance and best results is solved by the suggested hybrid TSCNDE model.

3.4 SUMMARY

The chapter introduces a novel approach to student performance prediction utilizing a DE coupled with a transient search capsule network. It encompasses various stages including data pre-processing, extraction of feature, selection of feature employing Fuzzy Equilibrium Optimizer (FEO), and feature classification. Additionally, the chapter delves into presenting the results and conducting discussions on the efficiency of the suggested model in expecting result of student.

CHAPTER 4

OPTIMIZED DEEP LEARNING BASED STUDENTS PERFORMANCE ANALYSIS BASED ON THE INFLUENCE OF SOCIAL MEDIA

This chapter introduces a deep learning approach that takes social media influence into account. A hybrid optimal long short-term memory system is suggested. The collaborative behaviour of the bay-winged predator and the capturability of the gannet in capturing prey are integrated in the design of the proposed hybrid optimization (Gannet Hunt) to find the optimal global solution with a quick convergence rate. To reduce the computational overhead in learning, the suggested Gannet Hunt algorithm is first used to identify the best features.

4.1 OVERVIEW

All around the world, people of all ages use social media networks extensively. With the growing number of users of online communities, social media is often characterized by its ability to show the relationships among users. Individuals of all ages, both men and women, share data on social media (SM) platforms such as plans, thoughts, films, and personal photos [60-62]. In addition, a significant amount of time is wasted by students on SM platforms such as X, Facebook, and other websites. Nonetheless, via the sharing of content, appropriate usage of social media enables students to progress their knowledge and social skills. A vast majority of students use social media in their daily lives, regardless of the positive or negative repercussions. As an additional benefit of gathering information in the form of a status update or content update factor, games, activities, entertainment, dramas, movies, musical concerts, and several other notifications about new brands or advertisements gathered from magazines, TV, or newspapers through Facebook, X, or other social sites are considered. A vast majority of students use social media in their daily lives, regardless of the positive or negative repercussions. As an additional benefit of gathering information in the form of a status update or content update factor, games, activities, entertainment, dramas, movies, musical concerts, and several other notifications about new brands or advertisements gathered from magazines, TV, or newspapers through Facebook, X, or other social sites are considered. By connecting to the internet through mobile devices, individuals of all ages or students can access social networks at any time and from any location. On the other hand, because social media is so simple and easy to use, frequent usage of it negatively impacts students' lives and taints society as a whole. The amount of time students spend on SM has an influence on their academic scores, even after accounting for their academic performance [63].

A contemporary area of research involves analysing student act based on educational data mining (EDM), which incomes into account multiple issues that have a non-linear impact on academic studies [64]. Using data mining methods, the student analysis's performance yields superior results when educational institutions are taken into account. Large datasets for student performance analysis are accessible, which has prompted academics to build a variety of performance prediction techniques [65–67]. When examining a student's academic learning process, an analysis of their performance is essential. This performance is gauged by their grade point average (GPA). Students' use of SM has a beneficial or bad impact on their academic attainment. Students who spend a large amount of time on SM and the internet don't focus on their academics, which has a negative impact on their academic achievement. Furthermore, because it is difficult to learn without accessing social media, the negative impacts of using it increase the likelihood of student dropouts [68, 69]. Therefore, the EDM approach's examination of student performance is crucial for reducing student dropout rates and boosting academic activities to raise grades. However, the analytics based on the EDM are inefficient and scalable in performance prediction, and they suffer from ambiguity in natural language processing [70, 71].

The educational institution's mission is to achieve academic objectives by offering high-quality instruction through behaviour analysis of the pupils. These days, advancements in information technology have a beneficial and negative impact on online education, which has an impact on academics. Students spend a large amount of time on SM sites and less time learning, which has an impact on their academic performance. In order to help students attain high academic standing, it is necessary to examine their performance in light of social media's influence. Although a few strategies for predicting student performance using ML have been established, the approach's effectiveness is still impacted by imprecise predictions and a failure to take into account crucial characteristics. Thus, this chapter introduces a deep learning approach that takes social media influence into account.

4.2 PROPOSED METHODOLOGY

Most age groups in the population use social media often, which has an effect on pupils' academic achievement. In addition to the benefits, it has a negative impact on education since it leads students to prioritize social media over academics and spend excessive amounts of time on it. Although a number of ML methods were developed to analyse student performance while taking social media into account, the performance is still constrained by the lack of data and the imprecise forecasts. As a result, a hybrid optimal DL algorithm that takes into account the inspiration of social media is presented for forecasting student performance. The procedure is presented in Fig. 4.1.

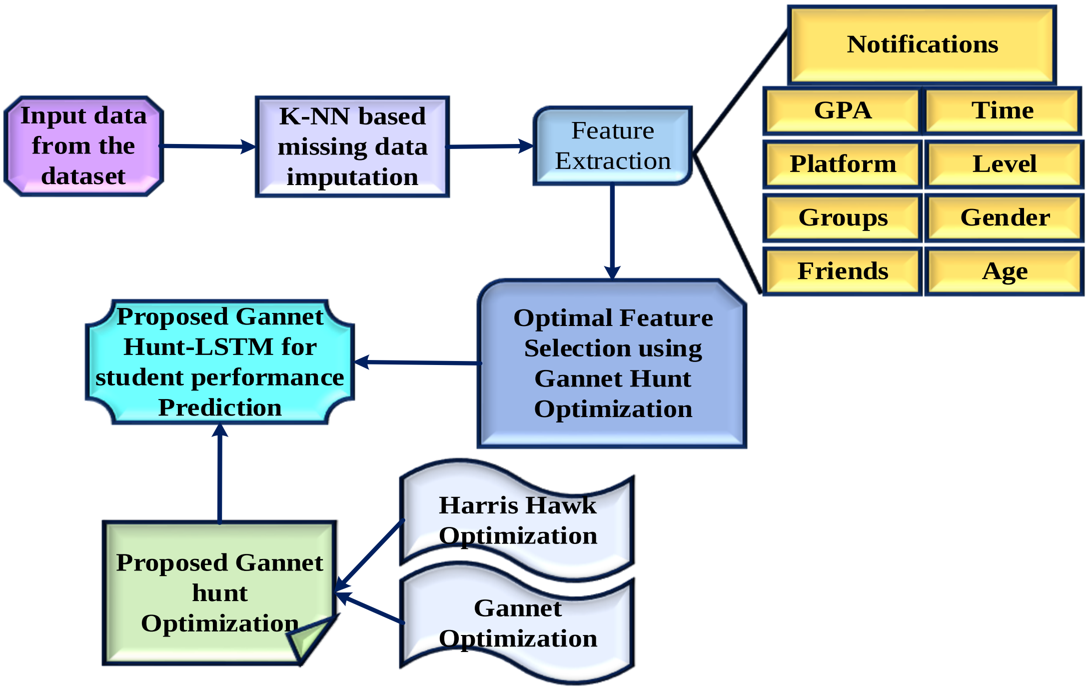


Fig. 4.1 Workflow of proposed Gannet Hunt-LSTM for student performance prediction

The provided workflow diagram outlines a comprehensive process for student performance prediction using the Gannet Hunt-LSTM model. Beginning with input data collection from a dataset, the process addresses missing data through K-NN based imputation, ensuring the dataset's completeness. Feature extraction follows, where various factors such as GPA, time spent, platform level, group involvement, gender, friendships, and age are extracted to capture diverse aspects influencing student performance. Subsequently, optimal feature selection is performed using a combination of optimization techniques including Gannet Hunt Optimization, Harris Hawk Optimization, and Proposed Gannet Hunt Optimization. These techniques aid in identifying the most pertinent features crucial for accurate prediction. Finally, the Gannet Hunt-LSTM model, incorporating both Gannet Hunt Optimization and LSTM architecture, is deployed for student performance forecast. This workflow provides a structured approach to leverage machine learning and optimization methodologies in educational settings, enabling informed decision-making and besieged interventions to enhance student outcomes.

The input data from Fig-share is used to forecast students' performance by taking social media influence into account. The most important features are then extracted to minimize computational overhead after the missing values in the dataset are imputed using the KNN method. The suggested Gannet Hunt procedure is then used to choice the best features from the recovered characteristics in order to increase forecast accuracy. In order to quickly discover the optimal global solution, the proposed Gannet Hunt algorithm combines the capturability of the Gannet with the hunting actions of the Harris Hawk. The LSTM is trained to make the generalization using the chosen optimal features. In order to reduce information loss and improve prediction accuracy, the LSTM's changeable parameters are adjusted using the suggested Gannet Hunt algorithm.

4.2.1 Data Acquisition

Data acquisition is a important process in various fields, encompassing the collection, recording, and retrieval of raw data from numerous sources for analysis and interpretation. This crucial step involves capturing information from sensors, instruments, databases, or external systems, conditional on the nature of the data being acquired. In fields such as scientific research, engineering, and industrial automation, data acquisition systems are utilized to gather data from physical phenomena, machinery, or environmental conditions. These systems typically involve sensors or transducers that convert physical signals into electrical signals, which are then digitized and stored for further processing. In modern data acquisition systems, digital interfaces and communication protocols are often employed to facilitate seamless integration with computers or other computing devices. The acquired data may include measurements of temperature, pressure, voltage, or other physical quantities, as well as textual, numerical, or multimedia data from various sources. Effective data acquisition ensures the availability of high-quality data for analysis, decision-making, and insights generation in diverse applications ranging from scientific research and industrial monitoring to healthcare and beyond.
The dataset made available on Figshare is the source of the input used to predict students' performance [72]. Through the use of an online survey, information about students' performance taking into account the impact of social media usage was acquired. Let the database is notated as Y and the total information in the database is specified as c. Then, the database is signified as,

$$Y = \{Y_1, Y_2, ...Y_c\} \tag{4.1}$$

Here, the K-NN technique is used to impute the missing values for the selected dataset in order to remove any biased student performance predictions.

4.2.2 K-NN based Missing data Imputation

KNN is a straightforward yet powerful model that is applied to regression and classification tasks in supervised learning. It functions according to the similarity principle, which conditions that a new data point's class is established by the classes of its closest neighbours in the characteristic space. Regression assigns the average mean of the rates of the KNN to the new data point, whereas classification assigns the majority class between the KNN. Since the approach is non-parametric, it doesn't make any assumptions on the information’s underlying distribution. Instead, it bases all of its predictions only on the training set. KNN is versatile and can handle complex decision boundaries, making it particularly useful for datasets with non-linear relationships. However, its performance may degrade with high-dimensional or noisy data, as it relies heavily on the local data structure. The value of k, which controls how many neighbours are taken into account while generating predictions, is one of the most important hyper parameters in KNNs. Selecting a suitable value for k is essential and frequently necessitates testing or cross-validation. The degree of similarity among data points is also measured using distance metrics, such as Manhattan distance or Euclidean distance. Even though KNN is straightforward, it's still a popular option for a diversity of classification and regression problems, mostly when interpretability and simplicity of implementation are important considerations. Due to its computing complexity which involves estimating the distances among each new info point and every point in the training set it could not scale well to huge datasets. Overall, KNN serves as a fundamental building block in machine learning and provides a valuable baseline for comparing the performance of more complex algorithms.

The biased result is derived from the performance forecast of the pupils whose data is lacking. Many studies used the deletion criterion to solve the problem, yet information loss and discontinuity still persist. Therefore, in order to prevent skewed results, the K-NN based methodology is used in the suggested method for the missing data imputation. Due to its effectiveness in managing missing values, numerous algorithms make extensive use of K-NN based imputation of the missing data. Several distance-based metrics, including the Euclidean, Hamming, Jaccard, Cosine, Manhattan, and so on, are used in the K-NN imputation of the missing information. The Euclidean distance measure is utilised to achieve the straightforward distance measure with highly intuitive similarity estimation. Therefore, in the K-NN based missing data imputation, the Euclidean distance based distance metric is used. The formula for the Euclidean distance measure is,

$$E_{dis}(v,u) = \sqrt{\sum_{d=1}^{o}\left(B_{rc} - B_{sc}\right)^2} \tag{4.2}$$

Where, the Euclidean distance between the two points (v,u) is designated as $E_{dis}(v,u)$, the feature that has the missing data is specified as B_{rc} and the feature that has the entire information is designated as B_{sc}, in which d refers to the total features. Here, the value of o needs to be chosen by estimating the Weight mean estimation and is signified as,

$$B_o = \frac{\sum_{r=1}^{R} b_R e_R}{\sum_{r=1}^{R} b_r} \tag{4.3}$$

Where, the nearest neighbors are designated as b_r , while the complete value of the missing feature is designated as e_r . Then, the expression for the nearest neighbor is formulated as,

$$b_r = \frac{1}{E_{(dis)^2}} \tag{4.4}$$

Next, use the observed values to fill in the missing values. The nearest neighbour is applied to fill in the loss of data in the dataset, completing it and preparing it for further processing.

4.2.3 Feature Extraction

For many ML and DL tasks, such as computer vision and NLP, feature extraction is an essential first step. It involves transforming raw input data, such as text, images, or signals, into a representation that captures relevant data for the task at hand. In the context of image processing, feature extraction aims to identify distinctive patterns or structures within the image that are essential for classification, detection, or other tasks. This process typically involves techniques such as edge detection, texture analysis, and key point detection, which help identify meaningful visual cues. In NLP, feature extraction involves changing text data into numerical demonstrations, such as word embedding's or bag-of-words vectors, to enable machine learning algorithms to process and analyse textual information effectively. Feature extraction plays a serious role in decreasing the dimensionality of the information while retaining vital information, thus facilitating extra efficient and effective machine learning model training and inference. The extracted features serve as inputs to downstream tasks such as clustering, classification, or regression, enabling the model to study from the information and make forecasts or choices founded on the learned representations.

The key characteristics are extracted from the dataset once all missing values have been imputed, simplifying processing and requiring less computational overhead. The dataset is used to extract features such as Grade Point Average (GPA), notifications, groups, friends, platforms used, amount of time spent on social media, educational level, student gender, and matching age groupings of the students.

4.2.4 Gannet Hunt based Optimal Feature Selection

Gannet Hunt based Optimal Feature Selection is a methodological approach rooted in bio-inspired computing, specifically inspired by the foraging behaviour of gannets, seabirds known for their efficient hunting techniques. In this context, the term "hunt" metaphorically represents the search for an optimal subset of features from a given set of information, akin to the bird's pursuit of prey. The process involves iteratively exploring the feature space to identify the greatest informative and discriminative features relevant to the task at hand, such as classification or regression. Drawing from the gannet's ability to identify and target specific prey items amidst a vast ocean, this approach aims to sift through a large pool of features to pinpoint those that contribute most significantly to the predictive performance of the method. Using the suggested Gannet Hunt method, the most important qualities are selected from the extracted features in the best possible way. Using the best feature selection criteria, the over fitting problem in forecasting students' performance is also resolved. The features with lower fitness values are removed in this case, and the best features with greater fitness are selected for additional processing.

Gannet Hunt Optimization: Gannet Hunt Optimization (GHO) is a metaheuristic optimization procedure inspired by the foraging performance of seabirds, particularly gannets, known for their efficient hunting strategies. In GHO, a population of candidate solutions, represented as potential prey locations, is iteratively updated to converge towards an optimal solution. The algorithm mimics the hunting behavior of gannets by employing a combination of exploration and exploitation strategies. During the exploration phase, candidate solutions explore the search space by randomly sampling potential prey locations. In contrast, the exploitation phase involves refining the search towards promising areas by adjusting the candidate solutions based on their fitness values. GHO utilizes a selection mechanism inspired by the success of gannets in locating and capturing prey efficiently. Through a process of natural selection and adaptation, candidate solutions evolve over iterations to improve their fitness and converge towards the optimal solution. The procedure's effectiveness lies in its aptitude to balance exploration and exploitation, leveraging both global and local search strategies to proficiently explore the search space and converge towards better quality solutions. GHO has demonstrated effectiveness in solving a wide range of optimization problems across various domains, creation it a respected tool in the field of optimization and computational intelligence. The combined hunting behaviour of the Hawk in capturing the quarry corresponding to the Harris Hawk Optimization [73] and the capturability of the Gannet in capturing the cunning fish corresponding to the Gannet optimization [74] are combined to obtain the global best solution with balanced intensification and diversification.

Motivation of Gannet Hunt Optimization: The Harris Hawk, a predatory bird with bay wings, hunts medium-sized rodents such as squirrels, rabbits, hares, and so on. Team hunting is thought to be the baywing hawk's clever method of bringing down its prey. The cooperative hunting approach has a higher success rate in obtaining the game than does hunting alone. Leapfrog movement, or cooperative hunting in packs, and foot-based quarry pursuit are taken into consideration for resolving the optimization problem. Nevertheless, they chase the prey with food for brief periods of time, but they do it frequently. The movement of the prey determines the baywing Hawk's hunting tactics. Several Hawks assault the target in a pack hunting situation from several angles in order to prevent them from fleeing. Similar to this, gannets are seabirds that pursue fish for sustenance. At a speed of 100 km/h, gannets plunge around 100 feet into the ocean. With binocular vision, the fish in the distance may be detected with more accuracy. As a result, the Gannet's rate of predation is greater. For this reason, to increase the convergence rate, the hunting behaviours of the bay-winged hawk and the gannet are integrated in the suggested Gannet Hunt optimization.

Mathematical Modeling: First, the extreme number of repetitions and the populace of the bay-winged predator in the search area are set for the proposed Gannet Hunt Optimization. Let the population size of the bay-winged predator is assigned with the variable E and the iteration with maximal count is indicated as $i_{\max}$. Next, the fitness measure is used to determine the matching solutions (positions) for the population of bay-winged predators that have been randomly located inside the search space.

Fitness Function: Fitness is a measure of how close the solution found by the bay-winged predators in the search space is to the target. The formula is used to measure fitness in order to select the best features.

$$\begin{aligned} SP_{fitness} = \sqrt{SP_{pre}(R)^2 + SP_{recall}(R)^2 + SP_{acc}(R)^2} - \\ \delta \times \frac{\alpha(R) \times \cos t(R)}{SP_{pre}(R) + SP_{recall}(R) + SP_{acc}(R) + 1} + \cos t_{\max} \end{aligned} \tag{4.5}$$

Where, the fitness function is designated as $SP_{fitness}$, the precision of the solution is designated as $SP_{pre}(R)$, the recall of the solution is indicated as $SP_{recall}(R)$, the accuracy of the solution is designated as $SP_{acc}(R)$. The adjustment coefficient is indicated as δ , which ranges among $[0,1]$. The dropped attribute is specified as $\alpha(R)$ and is assigned with the value of 0 and the selected attribute is assigned with the value of 1. The maximal upper bound cost is designated as $\cos t_{\max}$ and the sum of the cost corresponding to the feature sub-set is specified as $\cos t(R)$.

Searching for Quarry: The bay-winged predator waits for a long while many hours in the towering tree before spotting its target to catch. Because a predator's vision is greater than a human's, it can identify prey more precisely, even at a better distance. On the other hand, the resting strategy determines the diverse hunt for the quarry. When the bay winged predator rests near the family members and the quarry, then the condition for the resting strategy is indicated as $S \geq 0.5$. In contrast, when it rest on the random location then the condition for the resting strategy is indicated as $S < 0.5$. Hence, the solution of the bay winged Hawk based on the resting strategy is expressed as,

$$R(i+1) = \begin{cases} R_N(i) - V_1 |R_N(i) - 2V_2 R(i)| & S \geq 0.5 \\ (R_L(i) - R_o(i)) - V_3 - (D + V_4(A - D)) & S < 0.5 \end{cases} \tag{4.6}$$

Where, $R_N(i)$, $R_L(i)$and $R_o(i)$refers to the location of the randomly chosen predator, quarry, and mean position of the predators in the current iteration. The ranges of the random number like $V_1, V_2, V_3 \; and \; V_4$ are $[0,1]$. The boundary limit of the search area is designated as $A \; and \; D$ consistent to the upper and lower limits. The elevation of the diversification phase is employed to avoid the trapping at local optimal solution, which is accomplished by adjusting the scaling factor V_3. Here, the mean solution attained in the current iteration is accomplished through,

$$R_o(i) = \frac{1}{E} \sum_{y=1}^{E} R_y(i) \tag{4.7}$$

Where, E denotes to the total number of predators in search area.

Conversion from diversification to intensification: The transition from diversification to intensification represents a crucial shift in the exploration-exploitation trade-off. Initially, during the diversification phase, the algorithm focuses on exploring the search space widely to discover diverse solutions, aiming to uncover a broad range of potential solutions and avoid premature convergence to suboptimal solutions. This phase often involves strategies such as random exploration, mutation, or crossover operations in genetic algorithms. However, as the optimization process progresses and promising regions of the search space are identified, the procedure gradually shifts towards intensification. In this phase, the focus switches to exploiting the most promising areas of the search space to refine and improve solutions, aiming for convergence towards the optimal solution or local optima. The energy of the quarry determines how a phase changes from one to another. The quarry tries to flee by making haphazard leaps when it has enough energy to do so from the bay-winged predator. On the other hand, as energy levels drop, the quarry is unable to make haphazard leaps to get away from the predator, which causes the process of intensification to replace diversification. The energy of the quarry is obtained by,

$$Q = 2Q_0\left[1 - \frac{i}{i^{\max}}\right] \tag{4.8}$$

Where, the energy of the quarry is indicated as Q and the same at the initial stage is notated as Q_0. Here, the range of energy is between $[-1,1]$, wherein the range $[-1,0]$ depicts the lower energy and $[0,1]$ depicts the higher energy for escaping.

Capturing Quarry: Bay-winged Hawks locate the target during the searching phase, and then they use their wings to capture it during the intensity phase. Depending on the quarry's capacity for escape, the predator uses four different strategies to bring it to heel. The predator uses four different techniques to catch its prey: fine encircling, rough encircling, fine encircling with fast dives, and rough encircling with fast dives. The range of the escaping capability n is among $[0,1]$, wherein the value below 0.5 leads to the fine capturing and above 0.5 leads to the rough capturing by the predator.

(a) ***Fine encircling***: The assumption $n \geq 0.5\, and\, |Q| \geq 0.5$ is devised for the fine encircling of the bay winged predator. According to this criterion, the quarry is extremely energetic and wilfully leaps over the search area in order to escape the predator. The resolution reached at this point is shown as,

$$R(i+1) = \Delta R(i) - Q\left|CR_o(i) - R(i)\right| \tag{4.9}$$

$$\Delta R(i) = R_o(i) - R(i) \tag{4.10}$$

Where, the difference between the quarry and the solution accomplished by the bay winged predator during the current iteration is indicated as $\Delta R(i)$. The movement of the quarry C is signified as $C = 2(1 - V_5)$, wherein has the range of $[0,1]$.

(b) ***Rough encircling***: The assumption $n \geq 0.5\, and\, |Q| < 0.5$ is devised for the fine encircling of the bay winged predator. At this point, the quarry is less energy-hungry but more capable of fleeing. Hence, the bay-winged predator uses a cooperative attack to rough-encircle its prey. Several bay-winged predators target their prey in this instance, attacking from various angles to reduce their capacity to flee. At this point, the expression for updating the solution is shown below:

$$R(i+1) = R_o(i) - Q|\Delta R(i)| \tag{4.11}$$

By hybridizing the gannet's capturability, the bay-winged predator's ability to attack and capture its prey is improved. The gannet's acquired solution for catching its prey can be shown below:

$$R(i+1) = i * \lambda * (R(i) - R_{best}(i)) + R(i) \tag{4.12}$$

The factor λ depends on the capturability of the Gannet and is evaluated as,

$$\lambda = F * |R(i) - R_{best}(i)| \tag{4.13}$$

Where, the best Gannet in the current iteration is specified as $R_{best}(i)$, the capturability is specified as F. Then, the expression for the capturability of the Gannet is expressed as,

$$F = \frac{1}{X * i_2} \tag{4.14}$$

Where,

$$i_2 = 1 + \frac{i_1}{i_{\max}} \tag{4.15}$$

$$X = \frac{Y * l^2}{J} \tag{4.16}$$

$$J = 0.2 + (2 - 0.2) * V_6 \tag{4.17}$$

Here, the random number is specified as v_6 that has the range of $[0,1]$. The velocity and mass of the Gannet is designated as $Y\ and\ l$, which has the value of 1.5m/s and 2.5 kg. In this case, the quarry's potential to flee is diminished by the gannet's quick turn, which also captures the quarry to hasten convergence. Consequently, the outcome achieved by combining the gannet's capacity to be captured with the cooperative attack of a bay-winged predator is represented as,

$$R(i+1) = 0.5\left[R_L(i) - Q|\Delta R(i)|\right] + 0.5\left[i * \lambda * \left(R(i) - R_{best}(i)\right) + R(i)\right] \tag{4.18}$$

Here, based on equation (4.10), the rate of $\Delta R(i)$ is substituted and the equation becomes,

$$R(i+1) = 0.5\left[R_L(i) - Q|R_L(i) - R(i)|\right] + 0.5\left[i * \lambda * \left(r(i) - R_{best}(i)\right) + R(i)\right] \tag{4.19}$$

$$R(i+1) = 0.5R_L(i) - 0.5Q|R_L(i) - R(i)| + 0.5i\lambda\left(R(i) - R_{best}(i)\right) + 0.5R(i) \tag{4.20}$$

$$R(i+1) = \frac{1}{2}\left\{R_L(i)[1-Q] + R(i)[Q + i\lambda + 1] - i\lambda R_{best}(i)\right\} \tag{4.21}$$

Because of its improved capturing capability, the suggested hybrid expression used to capture the quarry achieves a fast convergence rate.

(c) ***Fine encircling with fast dives***: Fine encircling with fast dives is a strategy employed in optimization algorithms to efficiently explore the search space while converging towards the optimal solution. This technique involves first establishing a tight "encircling" region around promising solutions or areas of interest within the search space. The encircling phase aims to confine the search to a localized region where potential solutions are likely to be found. Once the encircling is achieved, the algorithm performs rapid "dives" or explorations within this confined region to thoroughly search for the finest possible solution. During the dive phase, the procedure intensively explores the local neighborhood of each candidate solution, employing techniques such as gradient descent or local search to fine-tune and refine the solutions. By combining the precision of encircling with the thoroughness of fast

dives, this strategy enables optimization algorithms to efficiently navigate complex search spaces, quickly identify promising regions, and converge towards high-quality solutions. Fine encircling with fast dives strikes a balance among exploration and exploitation, allowing algorithms to efficiently explore the search space while efficiently converging towards optimal or near-optimal solutions. The assumption $n < 0.5 \, and \, |Q| \geq 0.5$ is devised for the fine encircling with fast dives of the bay winged predator. The quarry has more vitality at this point, but there is less opportunity for escape. At this point, the bay-winged predator tries to catch the prey by making quick dives and precise encirclements. The suggested Gannet Hunt Optimization uses a guideline for the predator's subsequent move, which is stated as,

$$M = R_L(i) - Q|ER_L(i) - R(i)| \tag{4.22}$$

The bay-winged predator leaps with greater force to get away as it approaches its prey. As a result, the bay winged predator uses the quick dive mechanism based on levy flight during this stage. This mechanism is expressed as,

$$U = M + h \times W(G) \tag{4.23}$$

Where, the levy flight of the bay winged predated with the dimension G is referred as $W(G)$, and the random vector is with the size $[1 \times G]$ is notated as H. The solution accomplished by the bay winged predator is stated as,

$$R(i+1) = \begin{cases} M & if\ F(M) < F(R(i)) \\ U & if\ F(U) < F(R(i)) \end{cases} \tag{4.24}$$

(d) ***Rough encircling with fast dives***: Rough encircling with fast dives is a heuristic search strategy employed in optimization algorithms to efficiently explore and exploit the solution space. In this approach, the algorithm initially performs a rough encircling phase, where it broadly explores the search space to identify promising regions containing potential solutions. This phase aims to quickly encircle regions of interest without diving deeply into the search space. Once these regions are identified, the algorithm transitions into the fast dives phase, where it intensively explores the identified regions to refine and improve solutions. Unlike the initial rough encircling phase, the fast dives phase involves deeper exploration, focusing on exploiting the local structure of the search space to converge towards optimal or near-optimal solutions. By combining broad exploration with focused exploitation, rough encircling with fast dives strikes a balance among exploration and exploitation, allowing optimization algorithms to efficiently navigate complex

solution spaces and converge towards large-quality solutions in a timely manner. The assumption $n < 0.5\, and\, |Q| < 0.5$ is devised for the rough encircling with fast dives of the bay winged predator. The quarry is captured since it has little energy and is unable to flee during this period. The predator with bay wings' modified solution is stated as follows:

$$R(i+1) = \begin{cases} M & if\ F(M) < F(R(i)) \\ U & if\ F(U) < F(R(i)) \end{cases} \tag{4.25}$$

Where the rule used in the suggested Gannet Hunt Optimization for the predator's next move in this phase is expressed as follows:

$$M = R_L(i) - Q\left|ER_L(i) - R_o(i)\right| \tag{4.26}$$

Feasibility Check: Equation (4.6) defines fitness for all solutions, which is used to examine the feasibility of the updated solutions by the bay-winged predators.

Termination: The attainment of $i^{\max}$ or obtaining the optimal global finest explanation ends the iteration of the procedure. The pseudo-code is exposed in Algorithm 1.

Algorithm 1: Pseudo-code for Gannet Hunt Optimization

	Pseudo-code for Gannet Hunt Optimization
1	The number of iteration and population are initialized as: $i^{\max}$ *and E*
2	Locate the bay winged predator and quarry in search space
3	While $(i < i^{\max})$
4	Evaluate the fitness for all solutions
5	If $(Q \geq 1)$
6	Evaluate the solution accomplished at the searching for quarry phase
7	If ($n \geq 0.5\, and\, |Q| \geq 0.5$)
8	Update the solution accomplished at the fine encircling
9	Else if ($n \geq 0.5\, and\, |Q| < 0.5$)
10	Update the solution accomplished at the rough encircling

Else if ($n < 0.5\, and\, |Q| \geq 0.5$)
Update the solution accomplished at the fine encircling with fast dives
Else if ($n < 0.5\, and\, |Q| < 0.5$)
Update the solution accomplished at the rough encircling with fast dives
end
Update the best solution
$i = i++$
stop

Here, the best option, which is just the bay-winged Hawk's position, is used to choose the best features and train the LSTM to forecast the performance of the students. The balanced intensification and diversification should be a requirement of a competent meta-heuristic algorithm. The balanced searching methods are achieved in the suggested algorithm, whereby the intensification is increased by hybridizing the Gannet's capturability, and the diversification capability is strengthened by raising the scaling coefficient. As a result, the location determined by the fitness function yields the global best solution for resolving optimization problems.

4.2.5 Gannet Hunt-LSTM for student performance prediction

An example of a recurrent neural network (RNN) design called LSTM was created expressly to solve the vanishing gradient issue and identify long-term dependencies in data that is sequential. Unlike traditional RNNs, LSTM networks contain specialized memory cells that can preserve information over extended time intervals. Each LSTM cell is made up with the input gate, forget gate, and output gate. The cell state serves as a conveyor belt to transfer information across time steps. By selecting which data from the current input should be keep in the state of cell, the input gate regulates the flow of information into the cell state. The LSTM can discard obsolete or unnecessary data from earlier time steps by using the forget gate, which regulates the retention and removal of data from the cell state. Lastly, the instruction flow from the cell states to the LSTM cell's output is controlled by the output gate. These gates are organized by activation functions, typically sigmoid and hyperbolic tangent (tanh) functions, which determine how much information should be let through or discarded at each time step. As LSTM networks may choose store or discard information across extended sequences, they are especially useful for sequential data tasks including speech recognition, time series forecasting, and NLP. Because LSTM networks can manage sequential data and capture long-term dependencies, they are a powerful tool in many different domains that allow for more precise and reliable modeling of complex temporal interactions. Using the suggested Gannet Hunt-LSTM, the performance of the students is forecasted while taking into account the effect of their usage of SM. In this case, the LSTM uses the prediction and uses the suggested Gannet Hunt algorithm to optimize the modifiable parameter.

4.2.5.1 LSTM for predicting the student performance

Through regression and classification tasks, deep learning methods are used in many domains such as NLP, image processing, and other applications. RNNs are used to tackle the vanishing gradient problem that plagues CNNs and DNNs, two traditional DL techniques. RNN effectively manages the prediction of time series data by taking historical and present data into account. Furthermore, compared to a typical CNN, the prediction is more accurate due to the characteristics' dependence and the temporal data's consideration. RNNs process information using memory, but they are still hampered by the disappearing gradient problem and long-term dependences. Consequently, the vanishing gradient & long-term dependency concerns are resolved by LSTM. As a result, the recommended technique uses Gannet Hunt LSTM to predict the student's execution. Fig. 4.2 shows the design of the LSTM.

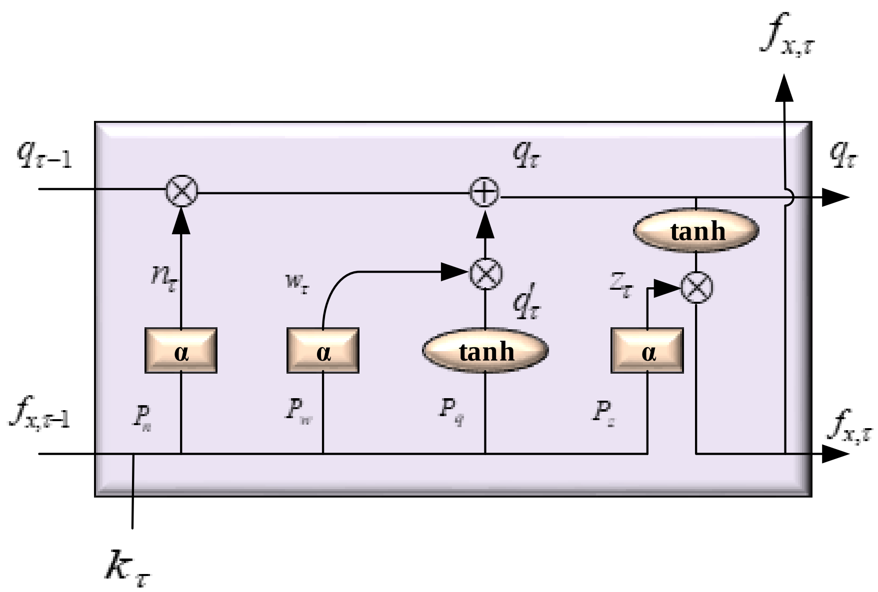


Fig. 4.2 Gannet Hunt LSTM

The optimal features extracted in the previous stage using the Gannet hunt are utilized by the LSTM for predicting the student's performance, which is indicated as k. LSTM comprises of various gates like input gate w_τ, forget gate n_τ, and output gate z_τ with certain functions of their own. Then, the cell state utilized for storing the long-term memory and is notated as q_τ. The hidden state f_τ is utilizes for carrying the information that accomplished at the previous stage of iteration and is utilized for estimating the present prediction. Here, the formulation utilized by the LSTM from the input k_τ to obtain the output f_τ is expressed as,

$$n_\tau = \alpha\left(P_n\left\lfloor f_{x,\tau-1}, k_\tau\right\rfloor + h_n\right) \tag{4.27}$$

$$w_\tau = \alpha\left(P_w\left\lfloor f_{x,\tau-1}, k_\tau\right\rfloor + h_w\right) \tag{4.28}$$

$$\overline{q_\tau} = \tanh\left(P_q\left[f_{x,\tau-1}, k_\tau\right] + h_q\right) \tag{4.29}$$

$$q_\tau = n_\tau \circ q_{\tau-1} + w_\tau \circ \overline{q_\tau} \tag{4.30}$$

$$z_\tau = \alpha\left(P_z\left\lfloor f_{x,\tau-1}, k_\tau\right\rfloor + h_z\right) \tag{4.31}$$

$$q_\tau = \tanh\left(q_\tau\right) \tag{4.32}$$

$$f_{x,\tau} = z_\tau \circ q_\tau \tag{4.33}$$

Where, the weights corresponding to the input, output, forget and cell state are notated as $P_w, P_z, P_n, and\ P_q$, and the biases are notated as $h_w, h_z, h_n, and\ h_q$ respectively. Besides, the sigmoid function is notated as α and the element wise multiplication is indicated as $\circ$. The LSTM takes into account both biases and weights as modifiable parameters to improve prediction accuracy. Here, the LSTM's parameters are changed using the suggested Gannet Hunt algorithm.

4.2.5.2 Learning Using Gannet Hunt Optimization

In order to expand the forecasting truthfulness of student performance, the weight initialization for the LSTM is important. Therefore, to optimize the LSTM weight adjustments and increase prediction accuracy, Gannet Hunt optimization is applied. Based on the mistake, the best learning strategy for the LSTM with Gannet Hunt is assessed in order to minimize information loss. Consequently, the MSE serves as the basis for the fitness used in the suggested Gannet Hunt optimization for learning the LSTM, which is expressed as,

$$Learning_{fitness} = \frac{1}{S_T}\sum_{x=1}^{S_T}\left(Out_O - Out_T\right)^2 \tag{4.34}$$

Where, the fitness utilized for learning the LSTM by the proposed Gannet Hunt algorithm is designated as $Learning_{fitness}$, the total samples is designated as S_T, the observed outcome is notated as Out_O and the target outcome is designated as Out_T.

As a result, data learning with the suggested Gannet Hunt reduces information loss and improves the LSTM's capacity for generalization, ensuring the best level of accuracy in student performance prediction.

4.3 RESULT AND DISCUSSION

This segment shows an examination of the newly introduced Gannet Hunt-LSTM for predicting student achievement while taking social media influence into account.

4.3.1 Experimental Setup

The analysis is assessed using the assessment measures, and the Gannet Hunt-LSTM is implemented using the PYTHON.

4.3.2 Dataset Description

The set of information used in the investigation of student performance taking social media influence into account comes from Figshare [72]. The UPSA uses an online survey to collect the dataset. A survey comprising 622 students was administered, taking into account several social media variables such as the amount of time consumed on SM, the total platforms used, and the number of friends in the groups, the educational grade, and several other factors.

4.3.3 Analysis Measures

To demonstrate the execution improvement, metrics through the performance parameters are assessed.

Accuracy: To assess prediction accuracy, the targeted outcome and the anticipated outcome of the Gannet Hunt-LSTM are compared. It is gauged by,

$$SP_{acc} = \frac{SP_{tp} + SP_{tn}}{SP_{tp} + SP_{tn} + SP_{fp} + SP_{fn}} \tag{4.35}$$

Where, the true positive is specified as SP_{tp}, true negative is designated as SP_{tn}, false positive is designated as SP_{fp} and false negative is specified as SP_{fn}, the accuracy is created as SP_{acc}.

F-Measure: This measure, which combines recall and precision using the harmonic mean, is stated as

$$SP_{FM} = 2 * \frac{SP_{pre} * SP_{recall}}{SP_{pre} + SP_{recall}} \tag{4.36}$$

Where, the recall is created as SP_{recall}, the precision is designated as SP_{pre}, and the F-Measure is specified as SP_{FM}.

Recall: Recall is measured by the accurate calculation, which is stated as, after taking into account false negatives and true positives.

$$SP_{recall} = \frac{SP_{tp}}{SP_{tp} + SP_{fn}} \tag{4.37}$$

Precision: Measured by precision, which is defined as the accurate prediction of student performance based on all positive classes,

$$SP_{pre} = \frac{SP_{tp}}{SP_{tp} + SP_{fp}} \tag{4.38}$$

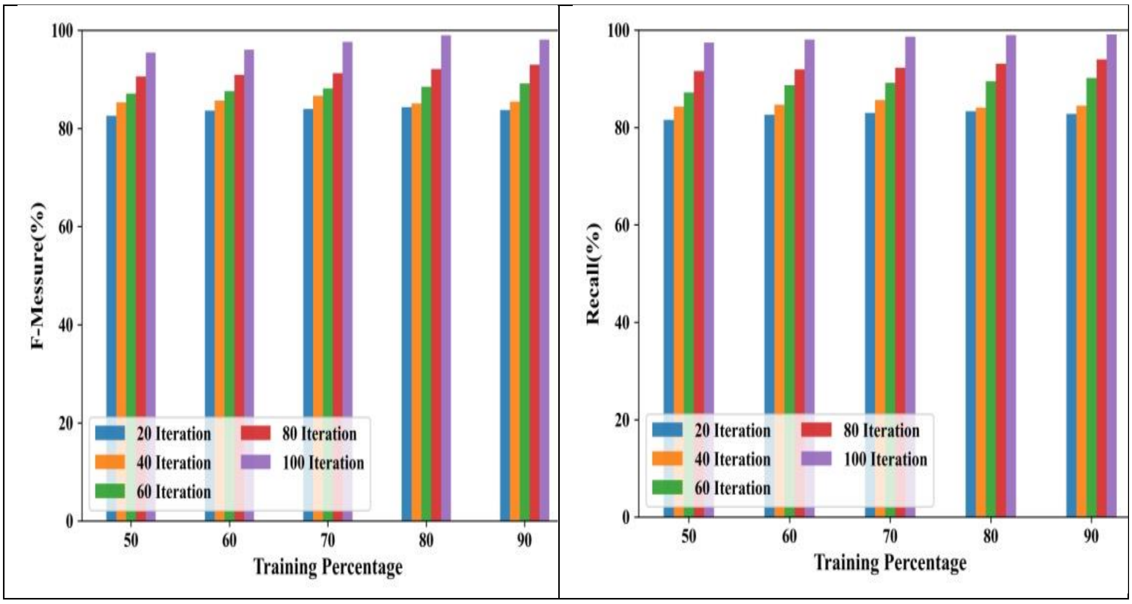
4.3.4 Analysis of Gannet Hunt-LSTM

This section goes into detail about the analysis of the newly introduced Gannet Hunt-LSTM based on the evaluation metrics, confusion matrix, accuracy, and loss analysis regarding the testing and training.

4.3.4.1 Analysis based on Evaluation Measures

Fig. 4.3 shows the Gannet Hunt-LSTM for forecasting student achievement by taking social media influence into account. Fig. 4.3(d) shows the accuracy estimate of the Gannet Hunt-LSTM, which is 83.78% with 90% of data learning and 20 iterations. With 90% of the data learned and 100 iterations, the same is 96.12%. Here, the accuracy study shows that the number of repetitions and the proportion of data learned have a significant influence on forecasting students' success when utilizing the suggested Gannet Hunt-LSTM. As a result, the study suggests that the approach performs better when there is more data for learning and iteration.

This is owing to the fact that learning with great data increases the LSTM's capacity for generalization, which raises prediction accuracy. In a similar vein, increasing the number of repetitions attempts to optimize the solution nearer the goal, guaranteeing a higher degree of performance prediction accuracy. Fig. 4.3(a), (b), (c) mention the analysis based on Recall, F-Measure, and Precision, respectively. In this case, the analysis by Recall, F-Measure, and Precision likewise offers the best results with higher iterations and data learning %. The study's iteration corresponds exactly to the iterations found in the suggested Gannet Hunt technique. The percentage of the dataset that was used to train the LSTM is shown by the data learning.



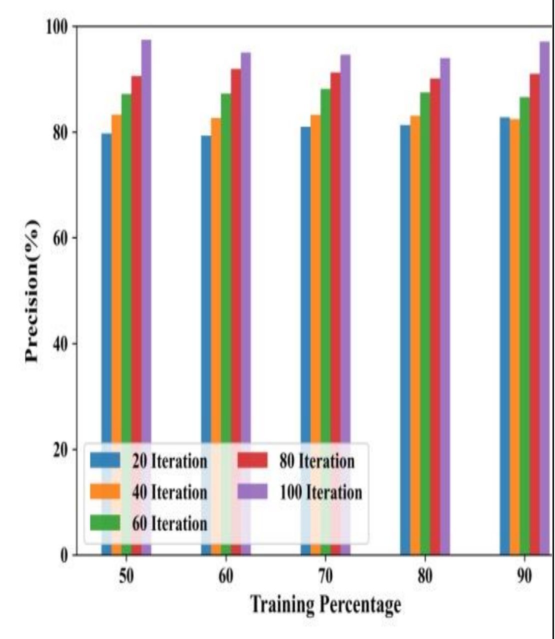
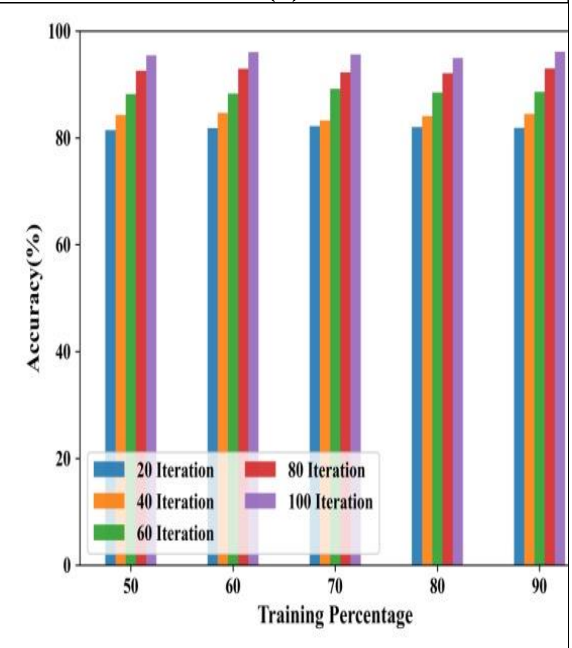
(a)	(b)
	
(c)	(d)

Fig. 4.3 Analysis of Gannet Hunt-LSTM in terms of (a) F-Measure, (b) Recall, (c) Precision and (d) Accuracy

4.3.4.2 Analysis based on Confusion Matrix

Fig. 4.4 displays the confusion matrix of the newly developed Gannet Hunt-LSTM for forecasting student performance by taking social media influence into account. The student performance in the classes A, B, B+, D+, D, and F was predicted by the suggested method with accuracy. Every time, the Class-C and Class-C+ predictions are made, the wrong one is made. As a result, the suggested Gannet Hunt-LSTM predicts more accurately.

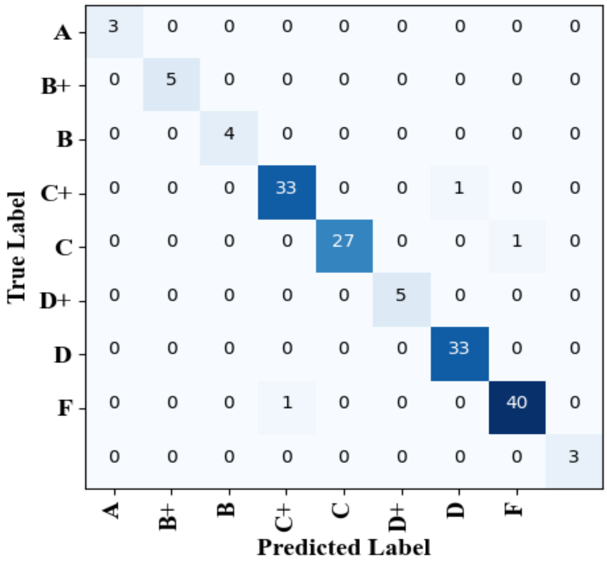
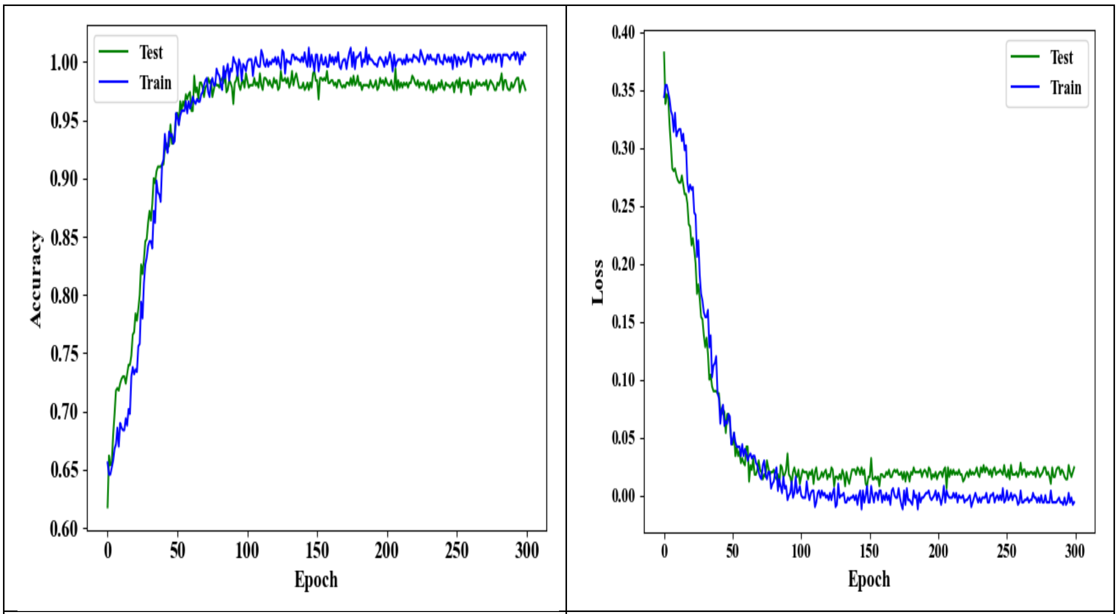


Fig. 4.4 Confusion Matrix of proposed Gannet Hunt-LSTM

Here, the academic GPA is used to evaluate the prediction of different classes, such as A, B, B+, C, C+, D, D+, & F. Students who achieve higher grades tend to have fewer friends on social media and use these sites for shorter periods of time. Likewise, students who use multiple platforms and friend groups receive more social media notifications. The pupil who received the best mark A student who has achieved a higher academic standing and spends minimal time on social media, has a minimum amount of friends, groups, and notifications on their accounts. On the other hand, the same is greater for pupils who received a F in their academic performance. Consequently, the analysis shows how social media negatively impacts students' educational experiences.

4.3.4.3 Testing and Training Accuracy and Loss Analysis

Fig. 4.5 displays the analysis of the Gannet Hunt-LSTM based on the training plus testing accuracy and loss. The analysis is based on altering the LSTM's period; a larger epoch reduces loss and improves forecast accuracy of student performance by taking social media influence into account.



(a) (b)

Fig. 4.5 Testing and Training Analysis in terms of (a) Accuracy and (b) Loss

4.3.5 Comparative Analysis

Traditional methods of predicting student performance, such as SVM [75], ANN [76], hybrid-CNN-RNN [77], and TSCNDE [73], are contrasted through the recently introduced Gannet Hunt-LSTM. Additionally, in order to illustrate the function of suggested optimization in suggested student performance prediction by taking social media influence into account, the performance of the suggested technique without optimization is also compared. Fig. 4.6 presented the comparison between 90% of the learning data and 10% of the testing data. Based on the presented outcomes, it can be inferred that the suggested methodology outperformed the conventional techniques. Figures 4.6(a), 4.6(b), 4.6(c), and 4.6(d) show the F-Measure, recall, precision, and accuracy analysis, respectively. In this case, every one of the four assessment metrics shows improved performance above the traditional approaches. The comparison approaches such as TSCNDE, LSTM, ANN, SVM, and CNN-RNN achieved 99.3%, 99%, 93.93%, 95.6169%, and 88.65%, respectively, while the Gannet Hunt-LSTM achieved an F-Measure of 99.9%. Similarly, every assessment measure yielded improved results.

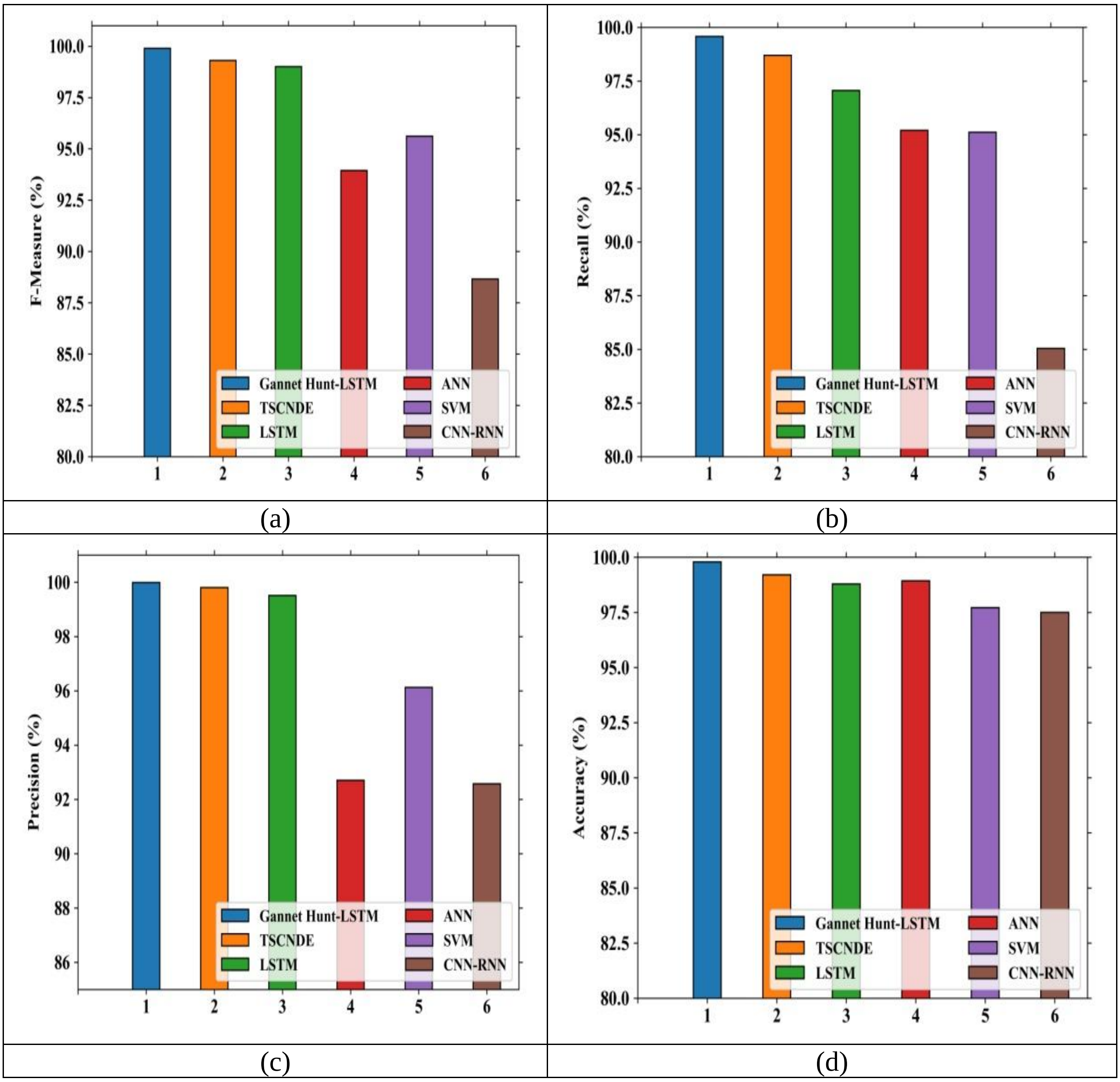


Fig. 4.6 Comparison in terms of (a) F-Measure, (b) Recall, (c) Precision and (d) Accuracy

4.3.6 Convergence Analysis

Fig. 4.7 compares the convergence of the suggested Gannet Hunt algorithm with the traditional Harris Hawk algorithm and the Gannet Optimization. The suggested algorithm's convergence is quicker than that of traditional algorithms, ensuring a lower time complexity.

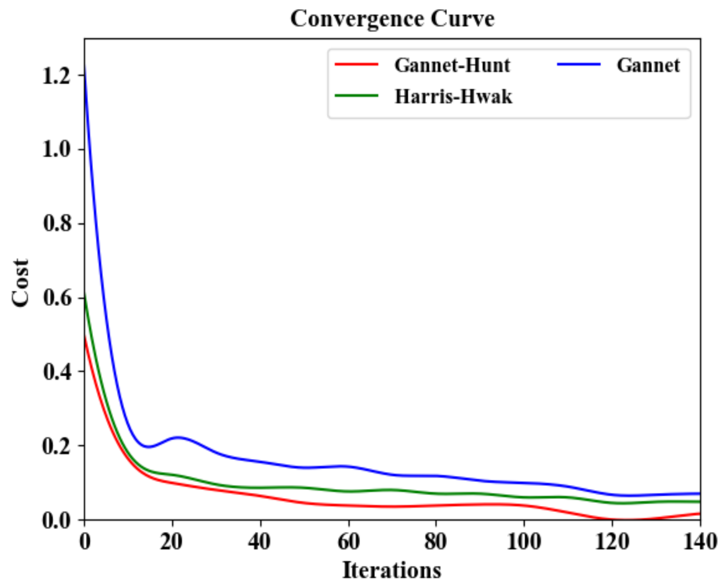


Fig. 4.7 Convergence Analysis

4.3.7 Discussion

Table 4.1 presents an examination of the Gannet Hunt-LSTM, which was introduced, along with its optimal results and traditional student performance improvement strategies, taking into account the impact of social media. Gannet Hunt-LSTM achieves a maximum accuracy of 99.78, which is 0.58%, 1.00%, 0.86%, 2.07%, and 2.30% better than traditional TSCNDE, LSTM, ANN, SVM, and CNN-RNN. The maximum precision achieved by Gannet Hunt-LSTM is 99.99, outperforming traditional TSCNDE, LSTM, ANN, SVM, and CNN-RNN by 0.19%, 0.48%, 7.29%, 3.87%, and 7.42%, respectively. With a maximum recall of 99.5738, Gannet Hunt-LSTM outperforms traditional TSCNDE, LSTM, ANN, SVM, and CNN-RNN by 0.88%, 2.52%, 4.39%, 4.48%, and 14.59%. With a maximum F-Measure of 99.9, Gannet Hunt-LSTM outperforms traditional TSCNDE, LSTM, ANN, SVM, and CNN-RNN by 0.60%, 0.90%, 5.98%, 4.29%, and 11.26%.

Table 4.1 Comparative Analysis

Methods/ Metrics	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F-Measure (%)**
Proposed Gannet Hunt-LSTM	**99.78**	**99.99**	**99.5738**	**99.9**
TSCNDE	99.2	99.8	98.7	99.3
LSTM (Proposed without optimization)	98.78	99.51	97.06	99
ANN	98.92	92.7	95.2049	93.93
SVM	97.71	96.12	95.114	95.6169
CNN-RNN	97.49	92.57	85.043	88.65

As a result, the analysis shows that the Gannet Hunt-LSTM that was introduced produced better results than the conventional approaches. To growth of the models rate of convergence, the hybridization of the capturability of the Gannet and the hunting behaviour of the Bay Wing Hawk is utilized. Additionally, the algorithm's balanced intensification and diversification capabilities aid in achieving the global best solution for choosing the best features and training the LSTM. In this case, the best feature selection lowers computational complexity, whereas the best LSTM learning minimizes information loss and improves the LSTM's capacity for generalization. As a result, the suggested Gannet Hunt-LSTM performed better.

4.4 SUMMARY

This study introduces the idea of predicting student performance while taking social media influence into account. A hybrid optimal LSTM network is suggested. The collaborative behaviour of the bay-winged predator and the capturability of the gannet in capturing prey are integrated in the design of the proposed hybrid optimization (Gannet Hunt) to find the optimal global solution with a quick convergence rate. To reduce the computational overhead in learning, the suggested Gannet Hunt procedure is first used to identify the best features. Additionally, by altering the adjustable parameters during the learning process, the Gannet Hunt algorithm's introduction lowers information loss and improves the LSTM's capacity for generalization.

CHAPTER 5

DISCRIMINABILITY ENHANCED TRANSFORMER ARCHITECTURE FOR STUDENTS PERFORMANCE PREDICTION USING ENHANCED FEATURES

This chapter explores the novel Discriminability Enhanced Transformer Architecture for Using Enhanced Features to calculate the performance of Student. It describes the Discriminability Enhanced Transformer Architecture in detail, the Distance-based Method that was used, and offers a perceptive analysis and interpretation of the outcomes. The effectiveness and promise of this unique approach in forecasting student achievement are carefully examined and analysed through careful experimentation and study, illuminating its significance in educational research and practice.

5.1 OVERVIEW

Data mining involves extracting information from extensive databases, especially those containing educational data [78, 79]. Educational data encompasses academic records, teacher and student details, circulars, syllabuses, and question papers [80]. The analysis of student performance is essential for educational institutions to ensure high-quality instruction and incentivize top achievers [81, 82]. ML algorithms are utilized to scrutinize enactment of students', including decision tree classifiers, gradient boosting, ada boosting, SVM, LR, NB, KNN, and RF [83]. However, these approaches have drawbacks such as lengthy classification times and a lack of linearity between dependent and independent variables [84]. Researchers have broadened their investigations to incorporate artificial intelligence techniques, such as, Levenberg-Marquardt algorithms, exercise-aware knowledge tracing, and enhanced RNN [85, 86]. These methods have boosted prediction accuracy to over 90% when combined with pre-processing strategies like filters [87]. Nonetheless, this methodology exhibits a lack of robustness for large sample sizes. This proposal advocates for a novel deep-learning approach to precisely ascertain student performance by inputting their historical performance and engagement data into artificial intelligence systems.

An essential tool for assessing students' comprehension and caliber of instruction is the prediction of their performance in academic settings. Historically, statistical methods have been employed; however, they are intricate, imprecise, and require a lot of time. On the other hand, challenges such challenging mistake detection, high data requirements, and growing temporal complexity confront machine learning. With the use of a DL model that can effectively predict the concert of student and a revolutionary outlier detection approach that speeds up processing of huge datasets, the study hopes to create a dependable technique for predicting student performance. An improved feature from the dataset is extracted in order to maximize the classifier's effectiveness. Statistical methods have traditionally been employed for gauging student performance, yet they are prone to errors and consume significant time. In response, machine learning methodologies are utilized, leveraging extensive datasets encompassing students' academic histories. Nonetheless, these approaches encounter hurdles such as susceptibility to errors, prolonged training times, and escalating temporal complexities. Consequently, scholars are turning to artificial intelligence methodologies, notably a hybrid deep learning algorithm, to attain precise and swifter identification of student academic performance.

One of the key challenges in virtual scholastic environments is to encourage active learning to improve students' ability to predict their academic performance. Monitoring changes in each student's knowledge levels through a series of instructional tasks is a common strategy for predicting student achievement, with a focus on exercise-related behaviors like accuracy and comprehension. Some student behavior variables appear to be equally useful in predicting student achievement, according to a few studies that have selectively focused on these attributes using subjective manual selection. To achieve highly accurate and faster identification of students' academic performance, a hybrid deep learning approach is employed. Initially, data from previous students is analyzed using an outlier identification technique, which detects outliers in large datasets that are sensitive to k-values and create local density issues through distance- and density-based methods such as k-nearest neighbor and reverse k-nearest neighbor. Subsequently, the multi-scale entropy approach is used to eliminate features from the data, and improved features are extracted from entropy-based features to enable reliable prediction, utilizing a hybrid deep learning method. A discriminability-enhanced transformer architecture is proposed to enhance feature-based prediction. The results section compares the suggested model with existing representations in terms of accuracy, f1-score, RMSE, precision, recall, and MSE demonstrating the suggested model achieves well than current models.

5.2 PROPOSED METHODOLOGY

Initially, an outlier identification technique is applied to the data of the former students. To ascertain the pace at which the directed density ratio varies among samples in order to decrease, KNN and reverse KNN are utilized. After the outlier has been eliminated, the multi-scale entropy technique is utilised to separate the data's features. A hybrid deep learning approach is used to filter enhanced structures from entropy-based information in order to provide an accurate forecast. To achieve better feature-based prediction performance, a transformer architecture with increased discriminability is created. The fundamental schematic of the suggested model is mention in **Fig. 5.1**.

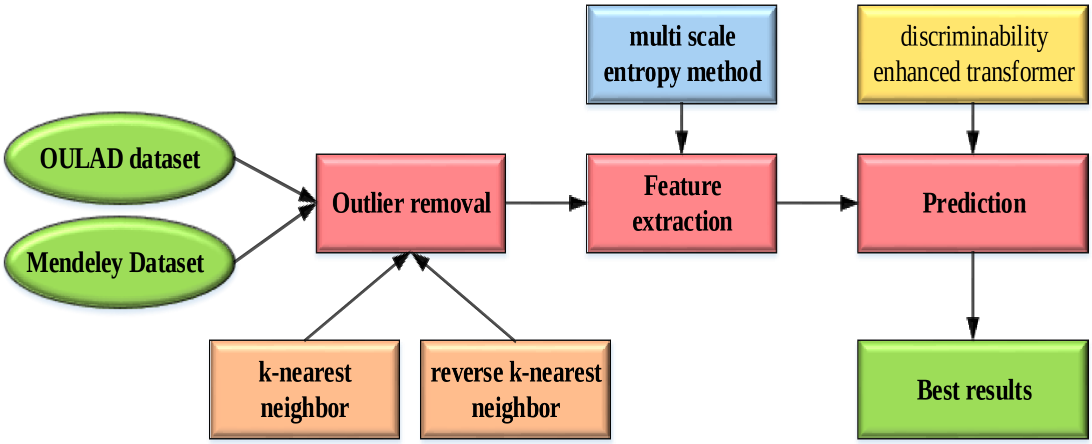


Fig. 5.1 Basic diagram of the proposed model

Outliers: An outlier could be defined as any information that is distinct, uncommon, inconsistent, untrue, or noisy. It follows that handling data and doing result analysis depend heavily on finding and removing outliers. High-dimensional techniques include those that are based on measurements of depth, density, statistics, distance, deviation, and so forth.

Distance based Method: The K-Nearest Neighbors (KNN) outlier identification method determines outlier scores by calculating the separation between every sample and the closest neighbors, recognizing that anomalies tend to be far from these neighbors. There are two distinct implementations of the KNN algorithm in terms of technical aspects. The average KNN calculates outlier scores based on the mean distance to a sample's k neighbors, while the maximum KNN uses the distance to the kth nearest neighbor. Mean-shift Outlier Detection (MOD) performs mean shift on the dataset by determining the average of a sample's k-nearest neighbors and replacing it with the original sample, determining outlier degrees by measuring shift distances after multiple iterations. Local gravity-based Outlier Detection (LGOD) calculates anomaly extent by adding gravitational values of samples based on their neighbors, considering it a distance-based technique due to its calculation involving the total of the distance norms and vector sums. Therefore, in cases where standard samples form clusters with varying densities, LGOD may overlook local outliers near dense standard clusters because a standard sample in a sparse cluster might resemble dense local outlier standard clusters more than normal samples.

Density based Method: Local Outlier Factor (LOF) addressed the aforementioned issues through a density-based technique. The outlier score computation involved dividing each sample's density ratio by its neighbor's local reachable distance. This computation was based on the principle that outliers have lower density compared to their neighbors, while typical samples have equal density to their neighbors. COF, based on the LOF paradigm, considers specific data distribution scenarios. By combining local reachable distance and kernel density effectively, LDF improved the estimation of local density. Consequently, LOF-based kernel density estimations were generated. INFLO broadened the collection of nearest neighbor model groups and computed density using both KNN and reverse KNN to enhance local density estimation. RDOS introduced the concept of joint KNN, deriving the last local density from kernel density estimates and the density ratio of the expanded closest neighbor set. This notion helps mitigate LOF's poor performance in complex data distribution circumstances. CELOF, another LOF enhancement technique, developed a new index estimation method by incorporating information entropy and constructing a new attainable distance feature discrimination methodology for authentic data retrieval. However, it may successfully identify a wide range of outliers in typical distribution settings. To illustrate this, consider embedding a 1D manifold in a 2D dataset; the density-based method misses outliers because sample density outside the 1-dimensional manifold is nearly equal to sample density inside the manifold. This study suggests an effective outlier identification technique by the rate of fluctuation of the focused density ratio addressing issues with various data distributions and the sensitivity measure k.

Calculation of Local kernel density: This paper examines the robustness and consistency of the density estimate when estimating kernel density and utilizing the extended closest neighbor method to get the local density of a trial point. The following equation provides the kernel density approximation for a specific model of the dataset $D = \{d_1, d_2, \ldots d_m\}$, where $d_p \in R^t$, $p = 1,23\ldots m$, t denotes the data's dimension:

$$b(d) = \frac{1}{m}\sum_{p=1}^{m}\frac{1}{y(d_p)^t} H\left(\frac{d - d_p}{y(d_p)}\right) \tag{5.1}$$

Where $H(\ )$ the kernel is function and $y(d_p)$ denotes the bandwidth of the kernel function at d_p. The Gaussian kernel function, which can be explained as follows, uses the kernel function.

$$H_{Gassian}(d) = \frac{1}{(2\pi)^t}\exp\left(-\frac{\|d\|^2}{2}\right) \tag{5.2}$$

Where the norm of d is denoted by $\|d\|$. In the implementation, the bandwidth y for the kernel function $y(d_p)$ is determined by computing the mean distance between each sample and its neighbour.

In order to more accurately gauge the density of the sample, the inverse and k-nearest neighbour functions are calculated. The $KNN(d_q)$ stands for d_q's closest neighbours. The samples d_p denoted by $RKNN(d_q)$ that have $\boldsymbol{xi}$ in their k-nearest neighbour set are the opposite of that sample's k-nearest neighbours. Consequently, the following formula yields the extended closest neighbour set of sample $d_q : EKNN(d_q) = RKNN(d_q) \cup KNN(d_q)$. The local density shown in d_q may be determined as follows using the extended closest neighbour set and Equation (5.2) in place of Equation (5.1):

$$b(d_q) = \frac{1}{j} \sum_{d_p \in EKNN(d_p)} \frac{1}{(2\pi)^t y^t} \exp\left(- \frac{\|d_q - d_p\|}{2y^2} \right) \tag{5.3}$$

Where the Euclidean distance between d_q and d_p is denoted by $\|d_q - d_p\|$.

The density ratio of samples is offered as a way to properly determine the degree of outliers in samples by combining density, direction, and distance to describe the association between a samples. The provided sample is provided as follows:

$$DRD(q,i) = \sum_{p=1}^{i} \frac{b(d_p)}{b(d_q)} t(d_q, d_p), d_p \in EKNN(d_q) \tag{5.4}$$

The formula for determining the sample local density $y(d_q)$ is expressed by Equation (5.3). In the feature space, the vector from d_q to d_p, denoted as $t(d_q, d_p)$, represents the distance and direction between them. The extended nearest neighbor set for d_q, denoted as $EKNN(d_q)$, and incorporates data on density, direction, and distance as described in Equation (5.4). The density ratio, $t(d_q, d_p)$, indicates the influence of d_p on d_q, with $\frac{b(d_p)}{y(d_q)}$ reflecting this relationship. A higher $b(d_q)$ compared to $y(d_q)$ suggest a stronger impact of d_p on d_q. $t(d_q, d_p)$ Contains information about both the distance and direction between points. The probability of d_q being an outlier increases with its distance from its nearest neighbors. In a normally distributed sample, $t(d_q, d_p)$ would exhibit inconsistent direction, resulting in minimal change in $DRD(q,i)$. Conversely, for a sample with an anomalous pattern, the neighbors would display a more irregular distribution, likely causing $t(d_q, d_p)$ to point in the same direction and leading to a significant alteration in $DRD(q,i)$.

According to the previously described research, the focused density ratio $DRD(q,i)$ represents the connection amid a sample and its adjacent samples. Furthermore, the degree of outlier in the sample is largely indicated by the change in $DRD(q,i)$. The throughout the formula can be used to govern the outlier degree of samples by figuring out how quickly the directed density ratio changes:

$$\Delta DRD(q,i) = \left\| \|DRD(q,i+1)\| - \|DRD(q,i)\| \right\| \tag{5.5}$$

The rate $\|DRD(q,i)\|$ is used to indicate the focused density ratio $DRD(q,i)$. The degree to which $\|DRD(q,i)\|$ is indicated as the number of nearest neighbours i increases is represented by $\Delta DRD(q,i)$. As such, the subsequent equation is used to calculate the final sample's outlier degree:

$$DCR = \sum_{i=1}^{I} \Delta DRD(q,i) \tag{5.6}$$

Here, I represents the synthetic hyper parameter and the sum of $\Delta DRD(q,i)$ for a range of i values. Where the cumulative changes in $\|DRD(q,i)\|$ are represented by DCR. The suggested model's pseudo code is represented by Algorithm 1.

Algorithm 1: Proposed outlier detection model

Input: dataset D, the neighbour size I
Output: Outlier score DCR
Initialization $DCR \leftarrow \{\ \}$;
Calculate the KNN graph of dataset D
For every $d_p \in D$ **do**
    Obtain $KNN(d_q)$ of d_q through KNN graph
    Obtain $KNN(d_q)$ of d_q through KNN graph

Get the extended neighbor set of d_q

$EKNN(d_q) = RKNN(d_q) \cup KNN(d_q)$

Evaluate local density $b(d_q)$;

End

For every $d_q \in D$ do

For i in 1 to $I-1$ do

Evaluate the directed density ratio $DRD(q,i)$ and $DRD(q,i+1)$;

Assess the directed density ratio's variation $\Delta DRD(q,i)$

End

Evaluate the final outlier score $DCR = \sum_{i=1}^{I} \Delta DRD(q,i)$;

End

Output the score of outlier DCR

Here's how the threshold T is calculated:

$$\begin{cases} \theta = median(K) + c \cdot MAD \\ MAD = e \cdot median(K - median(K)) \end{cases} \tag{5.7}$$

The item value in the middle is called the median, and K stands for the outlier score. e should ideally be 1.4826, and c should typically be set to 2.5. The k-nearest neighbour and reverse k-nearest neighbour algorithms are used to determine how quickly the ratio changes across samples in order to lower the directed density ratio. After the outlier has been eliminated, the multi-scale entropy approach is used to extract characteristics from the data.

Multi Scale Entropy Method: A approach for measuring the complexity and irregularity found in time series data across several temporal scales is the Multi-Scale Entropy (MSE) method. Fundamentally, MSE is figuring out a time series signal's sample entropy at different coarse-graining scales or levels. The probability that comparable data point patterns inside a certain window length will recur within the time series is measured by sample entropy. MSE provides insights into the complexity of the underlying dynamics by capturing variations in entropy over various time scales through a methodical increase in analysis size. Traditional linear methods may not fully represent the complexity of systems with non-linear and non-stationary tendencies, such physiological signals. In these cases, this method is especially helpful. The MSE approach is a useful tool in disciplines like physiology, neuroscience, and climate research because it provides a more thorough knowledge of the underlying structure and dynamics of complex systems by analysing entropy across several scales. Multi Scale Entropy (MSE) is the name given to a time series SE with many scales. If the entropy of a sequence falls monotonically as the scale factor rises and vice versa, the sequence's complexity is relatively simple. The steps involved in calculating MSE are as follows:

The scale factors b are obtained by the coarse granulation process. Following the coarse-graining series' setup with the first signal, the following outcome is provided:

$$H_q(b) = \frac{1}{b} \sum_{p=(q-1)b+1}^{qb} c_p, 1 \leq p \leq M/b \tag{5.8}$$

Where the factor that scales with an upward value is denoted by $b = 1$. When $b = 1$, the original set has not been the subject of coarse-grained analysis. The coarse-grained approach is based on the size of the window function or in the non-repetitive sliding averaging process.

For non-zero values of b, $\{c_p\}$ forms a coarse granulation series of length M/b. The final coarse-grained series is denoted by $\{H_q(b)\}$. At $b_{\max} = 18$, the highest point on the scale is configured. The MSE can be expressed as follows by using different scales $[1,2,\ldots,b_{\max}]$, the coarse-grain approach for multi-scale analysis, and the SE calculation of $\lfloor H_q(1), H_q(2), \ldots H_q(b_{\max}) \rfloor$.

$$MSE = \lfloor SE_1, SE_2 \ldots SE_{b_{\max}} \rfloor \tag{5.9}$$

The coarse graining approach can be used to obtain pattern data from the original time series corresponding to the scale factors by modifying the scale factors. To offer precise predictions, enhanced features are extracted from the entropy-based characteristics using a hybrid deep learning technique. It is suggested to use discriminability-enhanced transformer architecture to improve feature-based prediction.

Discriminability Enhanced Transformer Architecture: Contextual links and extra semantic information are typically associated with higher-level features. Low-level features, on the other hand, have greater spatial dimensions and retain finer, deeper characteristics like edges, corners, and lines. The DE unit aims to reduce the typical gap before integrating high-level and low-level features by improving the former's discriminability.

The humanoid visual model uses a variety of varying sized population receptive fields to highlight the retinal fovea since it is poor to even slight changes in location. This encourages us to use multi-field evaluating to boost high-level attribute discriminability and target object localization accuracy. On the other hand, it has been noted that conditions within a broad receptive field may enhance the interpretations of low-level data, hence minimizing noise and superfluous features. These two analytical vantage points informed the design of the DE unit, which addresses aspects at both the upper and lower levels. There are several branches in the module, and they all capture large-field contexts at various scales.

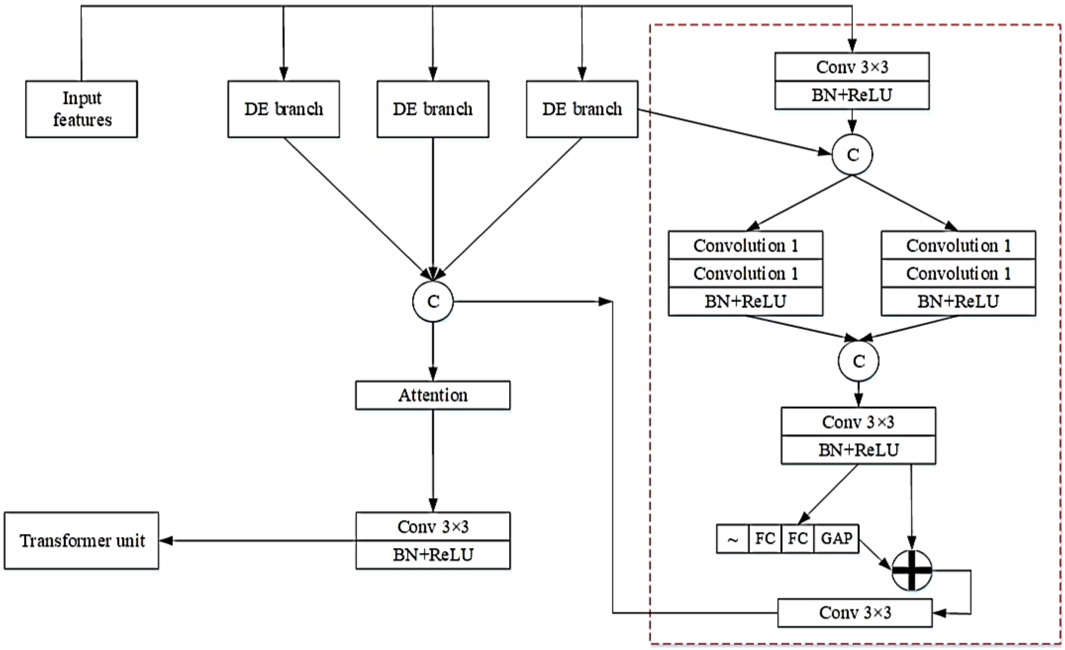


Fig. 5.2 Discriminability Enhanced model

The whole construction of the DE module is shown in **Fig. 5.2**. Assumed the incoming level-specific attributes, the DM unit's goal is to alter the features to make them more selective delegations. Exactly, the DE module analyses many fields simultaneously using four DE branches; the result of these branches is then grouped to provide the module's outcome. Each DE branch can be divided into three distinct stages of analysis: LFE, CFP, and LFF. The Transformer unit is added to improve the model. Recurrent models with good performance can handle sequence input in the form of $[k_1, k_2 \ldots k_i]$. Based on the input token p_i at location i and the prior hidden state H_{i-1}, they produce a classification of hidden states H_i lengthways the track of the effort order. The auto-regressive nature of the model means that the symbols it has generated so far are used as extra input for the next step. It is impossible to parallelize training samples because they are sequential by nature, which becomes crucial as sequence lengths rise. a creative layout that, by focusing just on attention strategies, demonstrates the global relationship among input and results. By entirely renouncing the conventional recurrent structure, the transformer does sequence input parallel processing. It is also challenging to represent the relationship between local characteristics globally because the transformer no longer employs the convolution function.

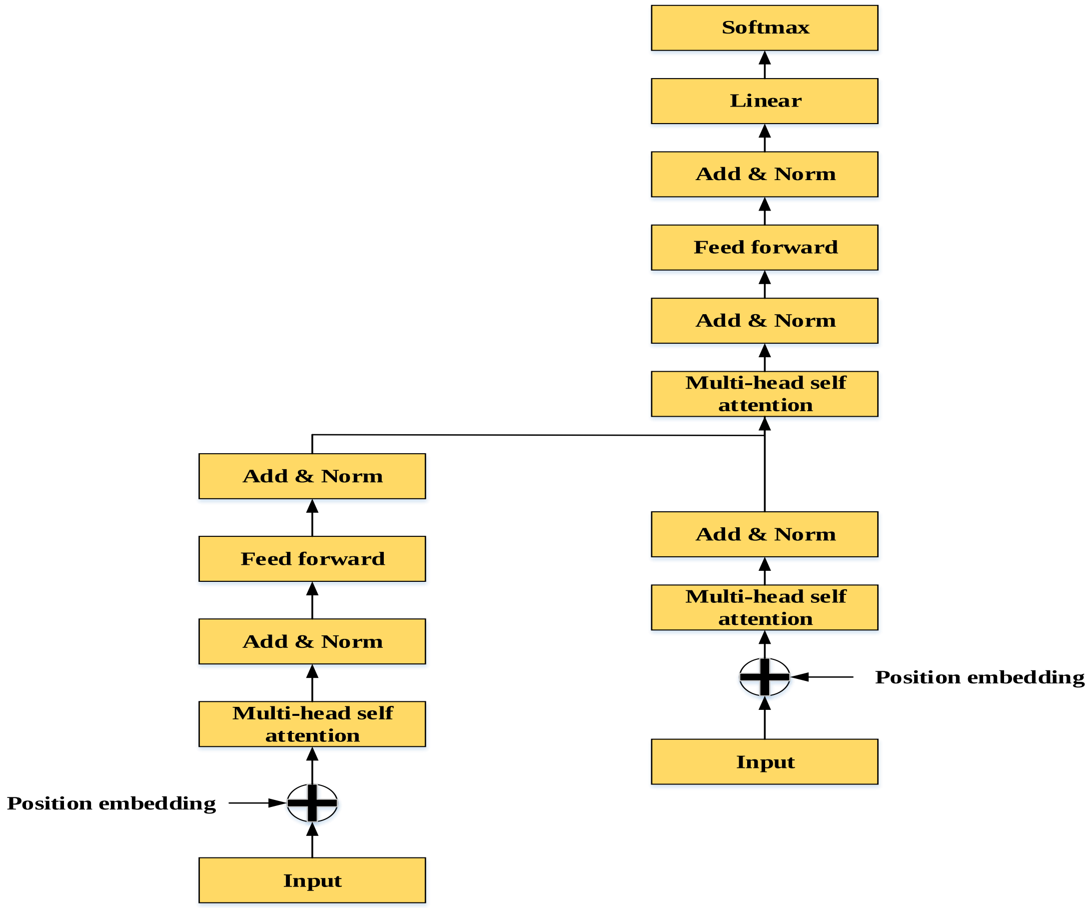


Fig. 5.3 Basic Architecture of Transformer

The basic form of Converter, which is a layered structure made by stacking Transformer bricks, is shown in **Fig. 5.3**. The Transformer block can be identified by its multi-head self-attention technique, residual connection, layer standardization module, and position-wise feed-forward network. The Transformer's input is often a tensor of type $R^Y \times R^s$, where s stands for classification length and Y for batch size. The input is first sent to an embedding layer, which generates a novel tensor $R^Y \times R^s \times R^D$. Each and every hot sign is changed to an embedded D dimension. After being added to a sinusoidal location encoding, the resultant tensor is then processed by the multi-head self-attention unit. The inputs and outputs of the multi-head self-attention are related via a residual link and a layer of normalization. A two-layer position-wise feed-forward network generates the combined output after linking the inputs along with a layer normalization layer and a lasting link. These sub layer connections that remain with layer normalization are said to take the following shape:

$$K_{out} = LayerNorm\left(M_F\left(K_{in}\right) + K_{in}\right) \tag{5.10}$$

Where M_F denotes multi-head self-attention based on location. The multi-head self-attention mechanism of transformer models sets them apart. This mechanism can be thought of as a process of learning alignment where each token in the pattern tries to learn from the others. Generally, a set of input embedding K of dimension $D_{\mathrm{mod}el}$ is transformed linearly to yield queries $Q_b = KW_b^p$ and keys $E_b = KW_b^e$ with dimension D_b, values $v_b = KW_b^v$ with size D_v. The dot products of all queries and keys are calculated by a single-head scaled dot-product attention, which then seperated each by the scaling factor $\sqrt{D_E}$ and applies a softmax function to get the mass of the rates.

$$C_b\left(Q_b, E_b, v_b\right) = soft\max\left[\frac{Q_b E_b^t}{\sqrt{D_E}}\right] v_b \tag{5.11}$$

Contrary to mapping with a single variety of linear changes, it is extra advantageous the H times to scheme the incoming tokens to several searches, keys, and values using dissimilar learned linear transformations. Thus, the notion of "multi-head self-attention" is introduced. Simultaneously self-attention calculation of all projected version queries, keys, and parameters yields H distinct outputs $Head_j$. Concatenating and projecting these $Head_j$ once again yields the multi-head self-attention.

$$C_H\left(K\right) = concat\left(Head_1 \ldots Head_H\right) W^L \tag{5.12}$$

Here, $$Head_j = C_H\left(KW_j^q, KW_j^e, KW_j^v\right) \quad 5.13)$$

In order to provide different keys, queries, and values, $W_j^q \in R^{D_{\mathrm{mod}el} \times D_e}$, and $W_j^e \in R^{D_{\mathrm{mod}el} \times D_e}$ denote the distinct and j^{th} variants of the linear projection on embedding's K, respectively. The symbol $W^L \in R^{H \cdot D_v \times D_{\mathrm{mod}el}}$ indicates the linear form projection on a point to the multi-head. Transformer, $D_E = D_v = D_{\mathrm{mod}el} / H$, here.

Next, the result of the multi-head self-attention unit is routed across two layers of feed-forward networking. "Position-wise" describes the independent actions of each point in this feed-forward layer.

$$PF(K_C) = \operatorname{Re}LU(0, K_C W_1 + Y_1)W_2 + Y_2 \tag{5.14}$$

Here, $W_2 \in R^{D_{model} \times D_{PF}}$, $W_1 \in R^{D_{model} \times D_{PF}}$ $B_1 \in R^{D_{PF}}$ and $B_2 \in R^{D_{D_{model}}}$ signify the biases and weights of the two layers, respectively.
A transformer is constructed by a series of stacked Transformer bricks. To obtain the combined result of their individual modules, position-wise feed-forward and multi-head self-attention modules located in a transformer block usually use layer standardization and a residual connection.

$$K_C = LayerNorm(C_H(K)K) \tag{5.15}$$

$$K_{PF} = LayerNorm(PF(K_C) + K_C) \tag{5.16}$$

Where the results of the multi-head self-attention module and the position-wise feed-forward unit are denoted by K_C and K_{PF}. Note that the stacked Transformer blocks have the same structure even though they might not have the same properties.

Understanding the differ in the applications of the Transformer block is essential. Transformers fall within one of three categories: Three methods are available: 1) encoder-only, 2) decoder-only, and 3) encoder-decoder. Figure 3 shows how a Seq2Seq problem is structured using an encoder-decoder structure by the vanilla Transformer for machine translation. Other from the ones listed for the encoder, other Transformer blocks are used by the vanilla Transformer decoder.

5.3 RESULT AND DISCUSSION

In order to develop and evaluate students' understanding and quality as well as to appraise the educational standards of the institution, it is imperative to forecast student performance in academic education. The strength of the suggested method was evaluated by contrasting it with a number of current techniques, such as LSTM, BiLSTM, CNN, and VI. The execution of the recommended strategy was assessed using a variety of performance metrics on datasets, including the Mendeley dataset [88] and the Open OULAD [89].

- ***Accuracy***

The previous method of performance evaluation employed the accuracy metric. Each match represents an exact prediction. The precision is calculated as the number of accurate forecasts separated by the total no. of inaccurate guesses.

$$Accuracy = \frac{No.\,of\ \ correct\ predictions}{Total\ \ No.\,of\ \ predictions} \tag{5.17}$$

- ***Precision***

Precision is ascertained by comparing True Positives (TP) to the mean of True Positives and False Positives (FP). Equation (5.18) represents the theoretically stated precision.

$$\Pr ecision = \frac{TP}{TP + FP} \tag{5.18}$$

- ***Recall***

Recall is defined as the ratio of True Positives (TP) to False Negatives (FN). The mathematical representation is expressed by equation (19).

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{5.19}$$

- ***F1 score***

It is calculated by adding the model's recall and precision scores. The accuracy statistic demonstrated how frequently a model generated an accurate forecast throughout the duration of the dataset. Recall and precision are occupied into account in F1. The process is expressed mathematically in equation (5.20).

$$F1 = \frac{2 * \Pr ecision * \mathrm{Re}\,call}{precision + \mathrm{Re}\,call} \tag{5.20}$$

- ***Mean Square Error***

The quality of a predictor is determined by its mean squared error (MSE), which is consistently positive. The MSE takes into account the prediction model's variance as well as bias, or how much the forecasts vary between data samples.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2 \tag{5.21}$$

Here, A_t is the observed value, n is the No. of samples, and F_t is the predicted value. The accuracy evaluation of the recommended approach on several datasets is mention in **Fig. 5.4**.

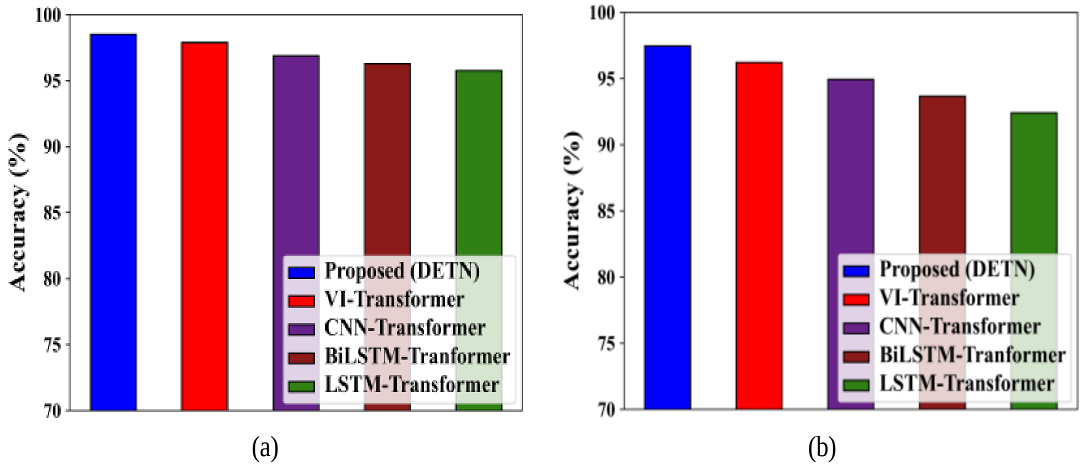


Fig. 5.4 Comparison of the suggested method's accuracy

Two distinct datasets are used in the suggested method's data collection. Fig. 5.4(b) correlates the accuracy rate of the suggested methodology with the OULAD dataset, while Fig. 5.4(a) displays the accuracy rate of the suggested approach using the Mendeley dataset. Four well-established methods are used by the suggested method to evaluate its performance rate. In the Mendeley dataset, the suggested technique achieved accuracy at a rate of 98.5%, while in the OULAD dataset, it achieved 97.47% accuracy. The evaluation revealed that the accuracy of the suggested approach was superior to that of other existing methods. The Mendeley dataset outperformed the other dataset in terms of % accuracy in forecasting student success.

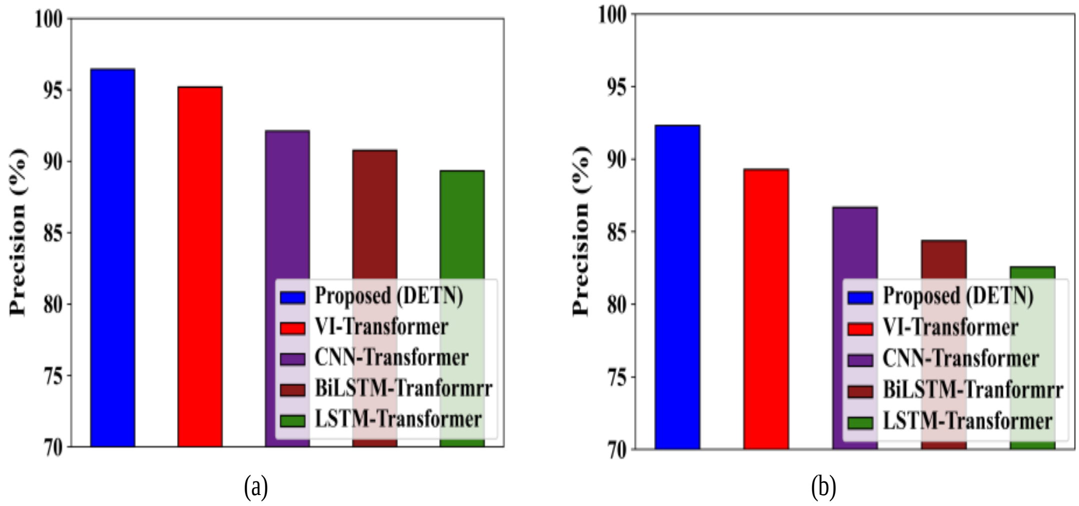


Fig. 5.5 Precision comparison of the proposed techniques

The precision rate comparing of the proposed method using various datasets is mention in **Fig. 5.5**. The strategy evaluates performance using four efficient current methods. The precision rate of the suggested technique is assessed using two distinct datasets. The precision score comparison of the recommended technique with the Mendeley dataset is mentioned in **Fig. 5.5**(a). The precision evaluation of the recommended scheme with the OULAD dataset is mention in **Fig. 5.5**(b). After going through a performance review process, the recommended approach obtained a rate of precision of 96.43% in the Mendeley dataset and 92.30% in the OULAD dataset. Thus, the suggested approach performs better than other previously used algorithms, which makes it more appropriate for the prediction sector. The recall rate and F1 score effectiveness of the suggested method are shown in

Fig. 5.6 for a variety of datasets.

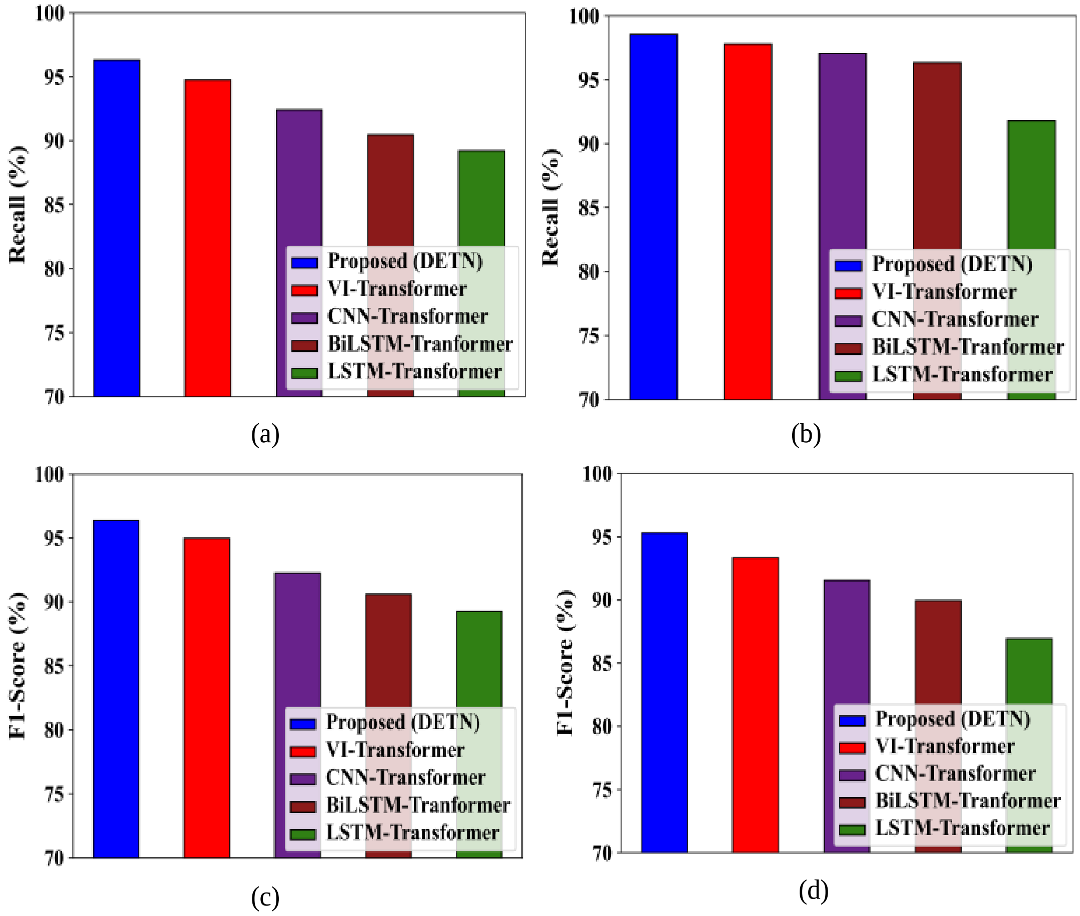


Fig. 5.6 Recall and F1 score comparison of the proposed method

To determine if the proposed method is beneficial in forecasting student performance, its recall and F1 score performance are analyzed. The execution of four other existing strategies is compared with the suggested method.

Fig. 5.6(b) compares the recall rates of the suggested strategy using the OULAD dataset, while

Fig. 5.6(a) displays the recall rate outcome of the suggested method using the Mendeley dataset. The comparison of the proposed technique's f1 score with the Mendeley dataset is displayed in

Fig. 5.6(c). The OULAD dataset is evaluated using the F1 score recommended by the suggested technique, as shown in

Fig. 5.6(d). Based on a performance review, the predicting performance of the recommended method was superior. The MAE and MSE comparison of the proposed method over several datasets is mention in **Fig. 5.7**.

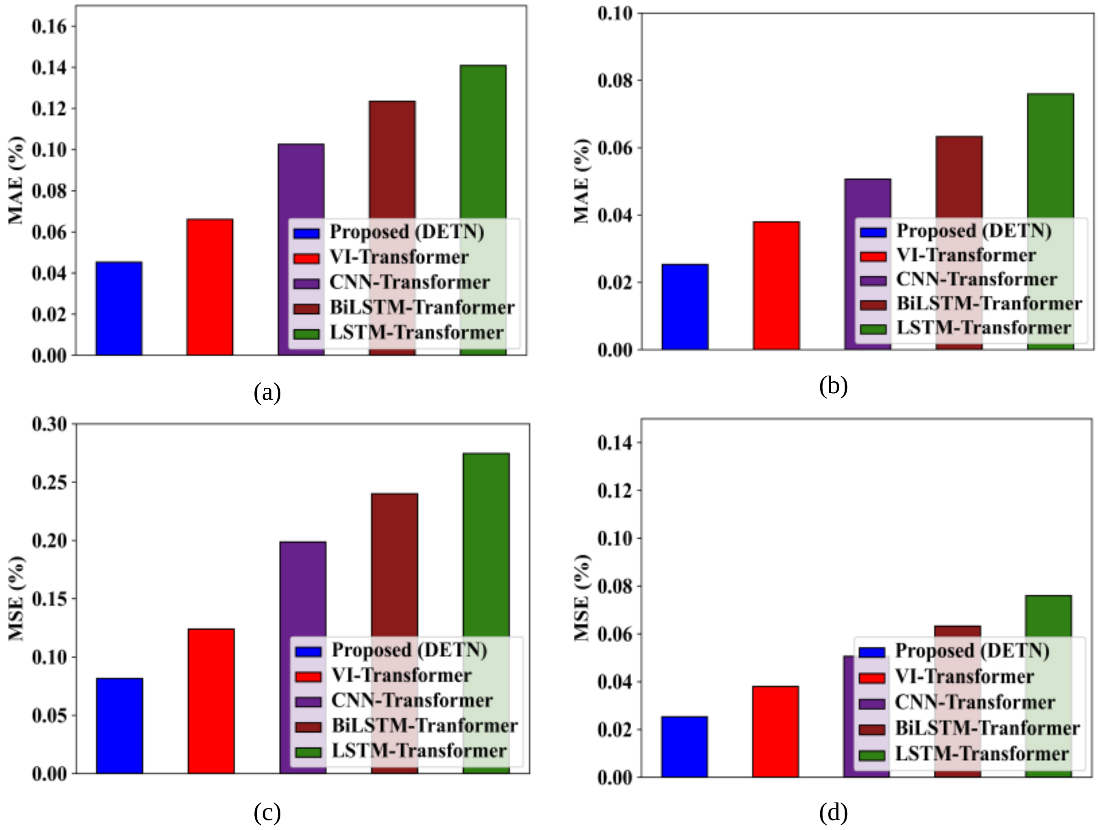


Fig. 5.7 MAE and MSE comparison of proposed method

Fig. 5.7 and the suggested error rate are contrasted. MAE and MSE, two distinct error evaluation matrices, are employed. The proposed method for assessing performance makes use of four different current methodologies. **Fig. 5.7**(a) and **Fig. 5.7**(b) show the MAE comparison of the recommended strategy utilizing the Mendeley dataset and the OULAD dataset, respectively, for the suggested method. **Fig. 5.7**(d) displays the MSE assessment of the recommended approach using the OULAD dataset, while **Fig. 5.7**(c) displays the MSE of the suggested method by the Mendeley dataset. According on the results of the strategy's evaluation process, the recommended strategy produced higher MSE and MAE rates. The processing times of the proposed technique are compared for various datasets in **Fig. 5.8**.

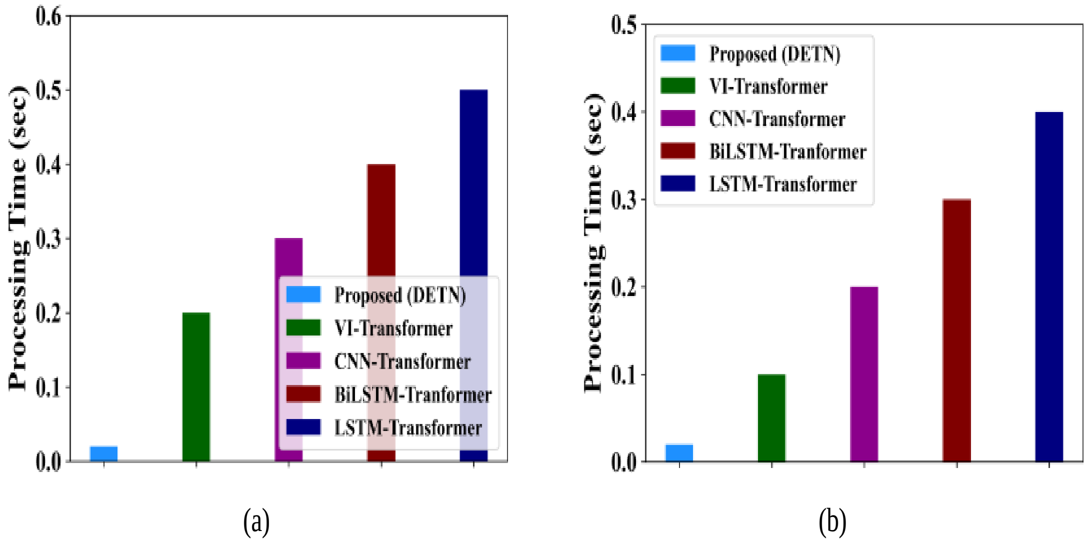
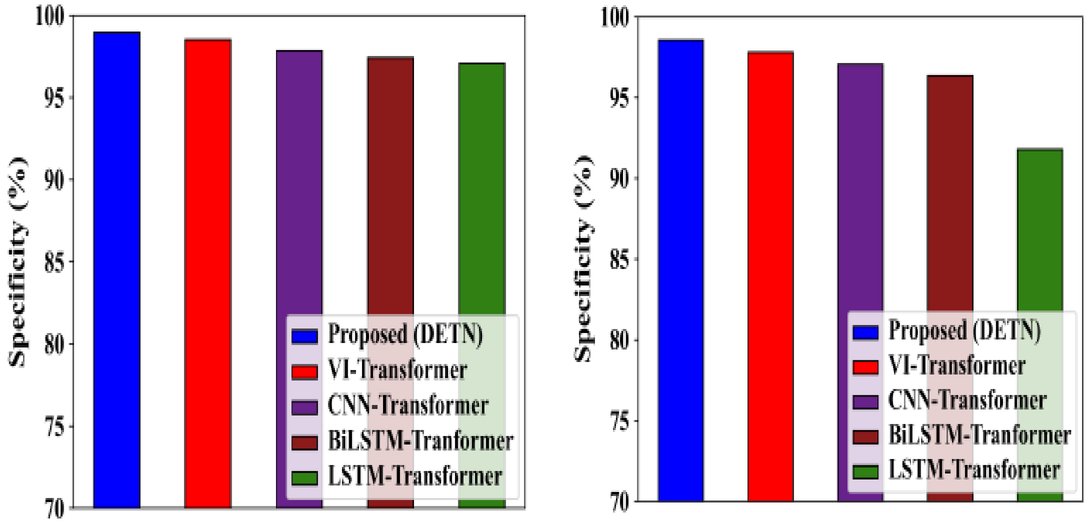


Fig. 5.8 Processing time evaluation of the proposed method

The suggested method uses two different datasets to calculate processing time. The processing time of the recommended method under the OULAD dataset is shown in **Fig. 5.8**(b), whereas the processing time of the recommended approach under the Mendeley dataset is shown in **Fig. 5.8**(a). According to the performance comparison, the recommended method produced faster processing times for both datasets. The suggested method takes 0.02 seconds to procedure in the Mendeley dataset and 0.02 seconds to procedure in the OULAD dataset. The suggested method took less time to process, based on the evaluation of processing time that was produced. As a result, the suggested approach fits the prediction process better. The processing times of the suggested and current models are mention in **Table 5.1**.

Table 5.1 Processing time of the proposed and existing models

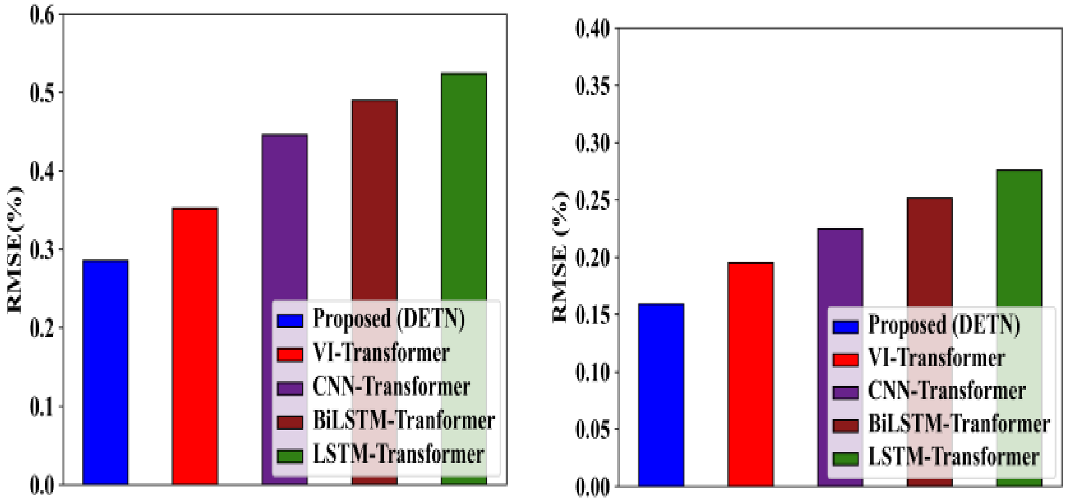
Dataset 1	
Proposed	**0.02**
VI-Transformer	0.2
CNN-Transformer	0.3
BiLSTM-Transformer	0.4
LSTM-Transformer	0.5
Dataset 2	
Proposed	**0.02**
CNN-Transformer	0.2
BiLSTM-Transformer	0.3
LSTM-Transformer	0.4
VI-Transformer	0.1



(a) (b)

Fig. 5.9 Comparison of the suggested method's specificity

The sensitivity rate of the recommended method is contrasted between several datasets in **Fig. 5.9**. The Mendeley dataset and a comparison of the specificity of the recommended strategy are shown in **Fig. 5.9**(a). The study revealed that the recommended strategy increased the Mendeley dataset's rate of specificity. Using the OULAD dataset, **Fig. 5.9**(b) compares the specificity rate of the recommended method. Consequently, the recommended approach yielded better forecast accuracy. The RMSE comparison for the recommended methodology is shown in **Fig. 5.10**.

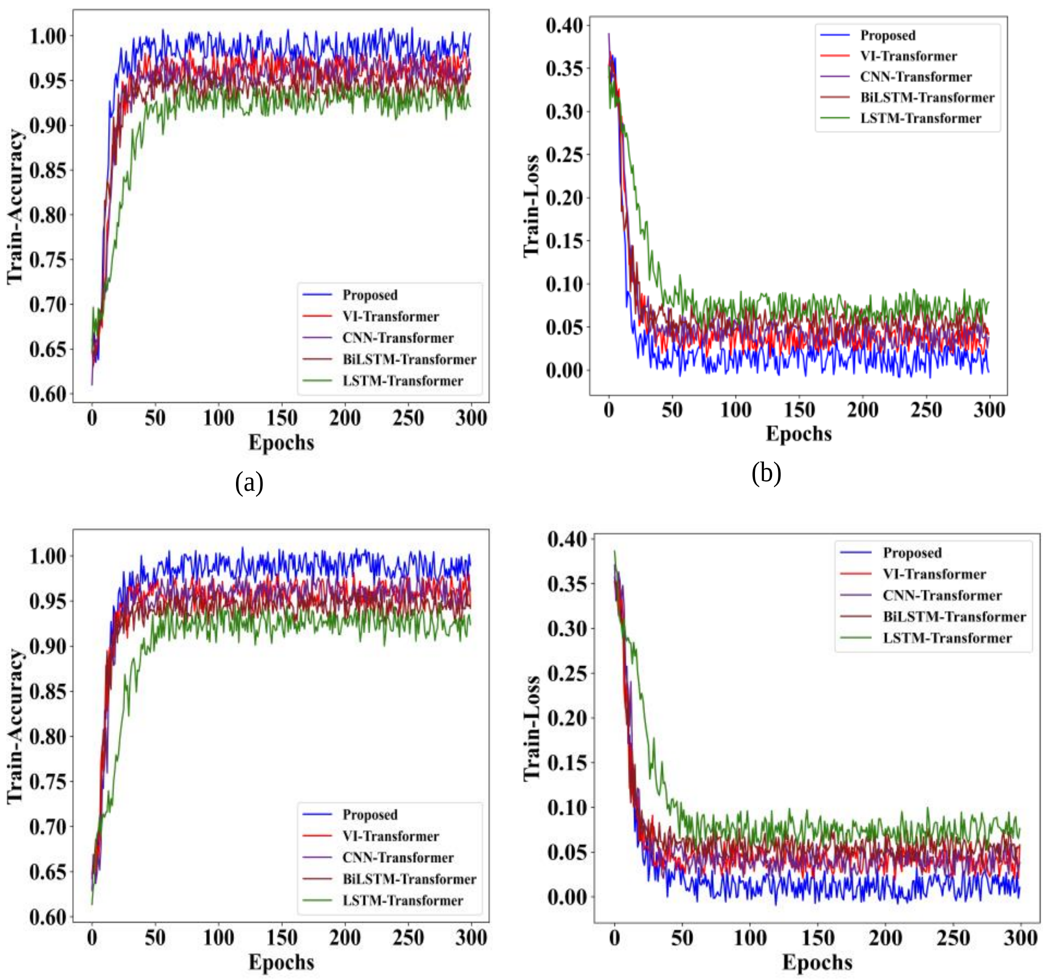


(a) (b)

Fig. 5.10 Comparison of the suggested method's RMSE

There have been four methods used to calculate the inaccuracy. As seen in **Fig. 5.10**(a), the suggested approach is contrasted using the Mendeley dataset to determine the RMSE. The suggested method for RMSE comparison using the OULAD dataset is shown in **Fig. 5.10**(b). In the resulting RMSE examination, the suggested solution fared better than existing methods, demonstrating its superiority as a predictor of student success.

The instructional accuracy and training loss comparisons for the recommended technique are mentioned in **Fig. 5.11**. In **Fig. 5.11**(a)–(b), the suggested training accuracy and loss within the Mendeley dataset are displayed. The training accuracy of the proposed technique with the OULAD dataset is mentioned in **Fig. 5.11**(c) and (d). The study revealed that the suggested approach produced less loss and a greater accuracy rate throughout the training phase. For predicting student achievement, the suggested approach is therefore more appropriate.



(a) (b) (c) (d)

Fig. 5.11 Training accuracy and training loss of the proposed method

Table 5.2 Detailed analysis of the suggested and current models' performances

Dataset 1					
Parameters	**Proposed performance**	**LSTM - Transformer Net**	**BiLSTM - Transformer Net**	**CNN -Transformer Net**	**VisionNet Transformer Net**
Accuracy	98.50	95.78	96.27	96.88	97.89
Specificity	98.94	97.07	97.40	97.83	98.50
precision	96.43	89.32	90.75	92.11	95.19
RMSE	28.56	52.39	48.99	44.57	35.19
recall	96.29	89.20	90.46	92.41	94.74
F1-score	96.36	89.26	90.61	92.26	94.96
MAE	0.045236	0.140836	0.123426	0.102596	0.066066
MSE	0.08161	0.274522	0.240012	0.198663	0.123892
Dataset 2					
Parameters	**Proposed performance**	**LSTM - Transformer Net**	**BiLSTM - Transformer Net**	**CNN -Transformer Net**	**VisionNet Transformer Net**
Accuracy	97.47	92.41	93.67	94.94	96.20
Specificity	98.52	91.77	96.32	97.05	97.79
precision	92.30	82.55	84.37	86.66	89.28
RMSE	15.91	27.55	25.15	22.50	19.48
recall	98.52	9177	96.32	97.05	97.79
MSE	0.025316	0.075949	0.063291	0.050633	0.037975
F1-score	95.31	86.92	89.95	91.56	93.34
MAE	0.025316	0.075949	0.063291	0.050633	0.037975

A thorough performance evaluation of the offered and existing models is given in **Table 5.2**. A comparison of several metrics, adding accuracy, recall, specificity, precision, MSE, F1-score, and MAE, is shown in it. The models' performances are compared on two datasets. The proposed performance in Dataset 1 achieves the highest accuracy, specificity, precision, recall, F1-score, and lowest values of RMSE, MAE, and MSE, consistently outperforming other models across most metrics. Notably, the VisionNet Transformer Net exhibits competitive performance, especially in accuracy and specificity. However, in Dataset 2, while the proposed performance remains strong, other models demonstrate improved performance compared to Dataset 1. Specifically, the LSTM-Transformer Net, BiLSTM-Transformer Net, and CNN-Transformer Net show substantial improvements in specificity, accuracy, F1-score, precision, and recall compared to Dataset 1. Nevertheless, the VisionNet Transformer Net still maintains competitive performance across most parameters. Overall, the results recommend that the projected technique consistently demonstrates superior performance across both datasets, with varying degrees of performance among other models depending on the dataset characteristics. Further analysis and experimentation may be necessary to understand the factors influencing model performance and optimize model selection for specific applications.

5.3.1 Discussion

Academic leadership relies on evaluations of students' progress throughout university courses to gauge their likelihood of success. Historically, teachers have gauged individual student performance through interactions such as in-class exercises and interim evaluations to pinpoint individuals who may falter or stop participating. In order to perform pre-processing, KNN doscover the k training set information points that are nearby to a new information point. Though KNN is straightforward, its performance can suffer due to dimensionality and k selection. Reverse KNN finds utility in anomaly detection and collaborative filtering. In the realm of time series data, Multi-Scale Entropy (MSE) effectively extracts features by gauging system complexity across various time scales, capturing insights overlooked by single-scale methods. Prediction is then carried out using a discriminability-enhanced transformer, a technique aimed at refining the temporal action localization capability of transformer-based models. Strategies to minimize rank loss and enhance learned representations' discriminability, such as employing a Scalable-Granularity Perception (SGP) layer and a Trident head for boundary models, are discussed. **Table 5.3** presents a comparison between existing and proposed techniques.

Table 5.3 Existing and proposed techniques comparison

Authors	Methods	Parameter
Ghassen Ben Brahim [90]	RF, SVM	Accuracy- 89%
Tsiakmaki et al. [91]	LSTM	Accuracy- 90%
Rivas et al. [92]	CNN	Accuracy- 92%
Poudyal et al. [93]	2D-CNN	Accuracy- 92%

Wu et al. [94]	LSTM-RNN	Accuracy- 96%
Proposed	Discriminability-enhanced transformer architecture	Accuracy- 98.5%

The table presents a comparison of existing and proposed techniques for predicting student performance, with a focus on the accuracy achieved by each method. Ghassen Ben Brahim utilized Random Forest (RF) and SVM algorithms, attaining an accuracy of 89%. Tsiakmaki et al. employed LSTM networks, reaching an accuracy of 90%. Rivas et al. utilized CNN, attaining an accuracy of 92%. Poudyal et al. applied 2D-CNN, also attaining an accuracy of 92%. Wu et al. utilized a combination of LSTM and RNN, reaching an accuracy of 96%. The proposed method employs a Discriminability-enhanced transformer architecture, achieving the highest accuracy of 98.5%.

This comparison highlights the gradual improvement in accuracy across different methodologies, with the proposed Discriminability-enhanced transformer architecture outperforming existing techniques by achieving the highest accuracy of 98.5%.

5.4 SUMMARY

The usefulness of the suggested hybrid deep learning approach the thesis concludes with a focus on predicting academic achievement in online learning contexts. The model obtains a 98.5% accuracy rate, which is higher than previous methods, with the use of discriminability-enhanced transformer architecture, feature extraction, and outlier detection. Subsequent investigations endeavor to delve deeper into datasets for amplified predictive precision and expand the approach for forecasting employee performance in organizational settings.

A hybrid deep learning technique is utilized to achieve a more precise and rapid assessment of students' academic performance in this scenario. Initially, an outlier detection approach is applied to previous students' data. Distance- and density-based methods are commonly employed for detecting outliers in large datasets, which can be influenced by k-values and local density issues. The pace at which samples differ in their directed density ratio is determined. The terms "k-nearest neighbor" and "reverse k-nearest neighbor" are utilized. After removing outliers, the multi-scale entropy technique is utilized to take characteristics out of the data. To achieve accurate prediction, enhanced features are then derived using entropy-based characteristics through a hybrid deep learning approach. The proposal suggests improving feature-based prediction by employing a discriminability augmented transformer architecture. Accuracy, f1-score, precision, recall, RMSE, and MSE are the metrics used in the results section to compare the suggested model with existing ones. The suggested model attained the highest accuracy range of 98.5% compared to other models. In the future, the research aims to explore additional datasets to accurately predict student performance and expand its scope to predict employee performance in organizations using more advanced methodologies.

CHAPTER 6

SHAPLEY EXPLAINABLE DEEP LEARNING BASED KNOWLEDGE DISTILLATION FRAMEWORK FOR STUDENT'S PERFORMANCE PREDICTION

This chapter suggest that predicting student academic performance on e-learning platforms is increasingly crucial in educational settings. Despite the recent advancements in Deep Learning (DL) technology, various methods have emerged with inherent limitations, including increased computational complexity and susceptibility to overfitting issues. Recognizing these challenges, a novel framework has been proposed to address these concerns and achieve more accurate student performance predictions with reduced computational complexity.

6.1 OVERVIEW

Unexpected developments in improved technology-based learning stands have resulted in a massive increase in data connected to online education, enabling the acquisition of several educational archives [96]. E-learning systems have been shown to offer a high rate of flexibility and engagement in recent years owing to their ease of use by learners on a variety of devices and at different times [97]. Mainly, such systems are primarily used by professors at hybrid and online colleges to collect and disseminate instructional materials to students [98]. Conversely, in-person teachers are able to learn about their students' behaviors and can more accurately anticipate their academic success [99]. While virtual communication between online instructors and students is possible, it might be challenging to evaluate each student's performance [100].

In recent years, educational institutions have increasingly transformed into highly competitive environments due to a growing interest in various courses and fields of study [101]. Additionally, the globalization of education has expanded opportunities for students to study abroad or pursue online courses, further intensifying competition among institutions to attract top talent [102]. Educational institutions are under pressure to continuously innovate and improve their programs, facilities, and services to remain competitive in attracting students and maintaining their reputation in the increasingly crowded landscape of higher education [103]. However, analyzing student academic efficacy in depth is a difficult task, because a student's academic execution may be impacted by a No. of social, economic, cultural, demographic, and academic background elements [104]. Developing an effective accepting of the factors influencing student academic execution requires a comprehensive analysis of data, which can be achieved through the utilization of ML and DL models. ML and DL techniques offer powerful tools for extracting insights from large and complex datasets, allowing educators and researchers to identify patterns, trends, and correlations that may affect student outcomes [105]. Through the application of ML and DL, educational stakeholders can gain actionable insights into the factors that impact student performance, leading to advance of targeted involvements and personalized support systems to enhance academic success for all students [106]. DL and ML are able to extract more meaningful characteristics from the input samples and make precise predictions about the results [107]. A few of the earlier techniques for assessing students' academic achievement were SVM [108], DNN [109], artificial neural networks (ANN) [110], etc.

6.2 PROPOSED METHODOLOGY

Predicting student academic performance on e-learning platforms is increasingly crucial in educational settings. Recognizing these challenges, a novel framework has been proposed to address these concerns and achieve more accurate student performance predictions with reduced computational complexity. This framework aims to streamline the prediction process by optimizing model architectures and refining training methodologies to mitigate overfitting risks. By carefully balancing model complexity and performance, the proposed approach seeks to enhance prediction accuracy while minimizing computational resource requirements. Through innovative techniques and algorithmic adjustments, the framework endeavors to provide educators and administrators with reliable insights into student performance on e-learning platforms, facilitating more informed decision-making and targeted interferences to support student success.

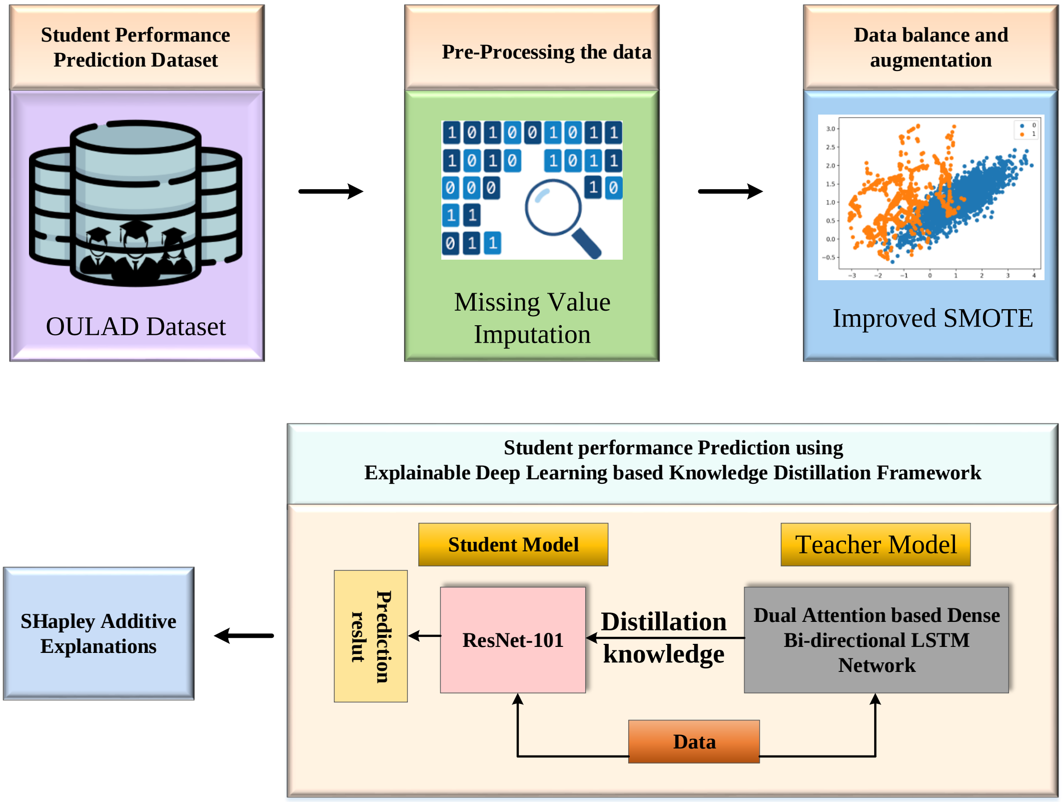


Fig. 6.1 Architecture of proposed framework

Improve quality of the dataset, the input data undergoes a series of pre-processing steps, beginning with Missing Value Imputation to address any gaps in the data. Following this, the dataset is subjected to data augmentation and balancing through the application of ISMOTE, enhancing the dataset's diversity and representativeness. The EDL_KDF framework is then introduced, comprising ResNet-101 as the student model and DAbDBi-LSTM as the instructor model, for predicting student performance. Finally, the framework employs SHAP analysis to assess the predicted outcomes, offering insights into the significance of each feature within the dataset and aiding in understanding of how these characteristics influence the prediction results.

6.2.1 Pre-processing the input data using Missing value imputation

Pre-processing the input data is a critical step in preparing data for ML models. It involves several procedures to clean, transform, and enhance the basic data so that it is suitable for model training and analysis. Proper data pre-processing can significantly impact model performance and reliability. The key steps in pre-processing the input data include:

- ***Data Cleaning:*** Contains identifying and managing missing, duplicate, or outlier values in the data. Missing values can be addressed through imputation (e.g., mean, median, or mode) or removal, while duplicate values and outliers may need to be removed or corrected.

- ***Data Transformation:*** This step may include scaling or normalizing the data to bring all features to a common scale. This is important for ML procedures that are iriitable to the scale of input features, such as neural networks and SVMs.

- ***Encoding Categorical Variables:*** Categorical variables are often encoded into numerical representations utilizing methods like one-hot encoding or label encoding. This allows models to handle categorical data effectively.

- ***Feature Engineering:*** This includes producing new features or transforming existing ones to advance model's ability to capture important patterns in the data. Techniques may comprise interaction terms, polynomial features, or domain-specific feature creation.

- ***Data Augmentation:*** In cases of imbalanced data, techniques like the ISMOTE can be used to create synthetic samples for the minority class, helping to balance the set of information.

- ***Dimensionality Reduction:*** Sometimes, reducing the number of features can improve model performance and reduce overfitting. Procedures like PCA or feature selection can be used to achieve this.

- ***Splitting Data:*** The data is typically divided into training, validation, as well as testing sets to facilitate model evaluation and hyperparameter tuning.

In order to handle the few missing values in the dataset, which typically arise due to various issues, one effective approach is to exchange them with mean value of the nearby non-missing values. This method involves calculating the average of the known values surrounding the missing data points. By estimating the missing values based on the mean of their neighboring data, this approach maintains the integrity and consistency of the dataset, ensuring that any analysis or modeling conducted on the data is based on a more complete and accurate dataset. This strategy helps in mitigating the impact of missing data on subsequent modeling tasks.

$$U_i = \frac{\mathrm{U}_{\mathrm{i-1}} + U_{i+1}}{b} \qquad i \in D \tag{6.1}$$

Where, missing value is signified as U_i, previous value of the missing value is signified as U_{i-1} like that, the next value followed by the missing value is denoted as U_{i+1}. The symbol for the natural number is $D = 1,2,3,\ldots$. By using this strategy, the dataset's missing values are successfully restored, improving the input data's quality.

6.2.2 Data balance using Improved Synthetic Minority Oversampling Technique (ISMOTE)

ISMOTE is an advanced version of the original SMOTE. Both techniques aim to address class imbalance in datasets, which can cause issues in machine learning models, such as biased predictions and reduced performance. ISMOTE enhances the original SMOTE approach by generating synthetic samples more effectively and intelligently, focusing on areas of the feature space where the data is sparse or underrepresented. In ISMOTE, instead of randomly selecting neighbors for generating new samples, the method uses advanced strategies to identify the most relevant or closest neighbors based on the distribution of the data. This targeted approach helps to create more realistic and diverse synthetic samples, better representing the minority class and reducing the risk of overfitting. By improving the quality and representativeness of the augmented data, ISMOTE helps improve performance and reliability of ML models trained on imbalanced datasets.

In addition to precisely defining the boundaries, ISMOTE aims to provide distinct synthetic samples for the minority class of unbalanced datasets to data augmentation through SMOTE generalization. The ISMOTE process is described as follows. First, a synthetic instance is made by using the SMOTE procedures, which are stated as follows.

$$Z = 2 * (v - u) + u \tag{5.2}$$

Where, number of initial synthetic instance is represented as Z, minority and majority class samples counts are denoted as v and u respectively. In second stage, remove the synthetic samples that are closer to the majority class compared to the minority, and the generated instances that are nearer the border created by SMOTE. The ISMOTE step by step process are given as below. Consider the new synthetic instance set as $\hat{x} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_Z\}$ and $k-th$ attribute value of the $\hat{x}_l$, $k \in [1, C]$ is denoted as $\hat{x}_l^{(k)}$, which are determine the accept or reject the instance created by SMOTE. Let set of majority and minority samples are denoted as $F_p = \{F_{p1}, F_{p2}, \ldots, F_{pm}\}$ and $F_q = \{F_{q1}, F_{q2}, \ldots, F_{qn}\}$ respectively. To decide whether to accept or reject synthetic instance, the distance among the E_{pm} and $\hat{y}_l$ moreover distance among the F_{qn} and $\hat{x}_l$ are estimated in initial step.

$$C_n(\hat{x}_l, F_{qn}) = \sum_{k=1}^{C} \sqrt{\left(\hat{x}_l^{(k)} - F_{qn}^{(k)}\right)^2}, \; n \in [1, u] \tag{5.3}$$

$$C_m(\hat{x}_l, F_{pm}) = \sum_{k=1}^{C} \sqrt{\left(\hat{x}_l^{(k)} - F_{pm}^{(k)}\right)^2}, \; m \in [1, v] \tag{5.4}$$

Two arrays were constructed using the equation above, and the results are shown below.

$$G_{\min} = \left(C_n(\hat{x}_l, F_{q1}), \ldots, C_n(\hat{x}_{l\,l}, F_{qn})\right) \tag{5.5}$$

$$G_{\max} = \left(C_m(\hat{x}_l, F_{p1}), \ldots, C_m(\hat{x}_l, F_{pm})\right) \tag{5.6}$$

Select the lowest value from the minority and majority arrays. The synthetic sample is either approved or rejected based on whether the minority value is less than the majority minimum value. That increase the speed of technique and increase performance. In next step, the noise is removed from the accepted synthetic instance $\hat{F} = \{\hat{F}_1, \hat{F}_2, \ldots, F_z\}$.The step computed the distance among $\hat{F}_l$ with all real minority F_q that given as below.

$$H_n(\hat{F}_l, F_q) = \sum_{j=1}^{v} \sum_{k=1}^{C} \sqrt{\left(\hat{F}_l^{(k)} - F_{qn}^{(k)}\right)^2} \tag{5.7}$$

Through the above expression the new t sample set via eliminate the far away samples from the minority is created as follow.

$$t = \sum_{i=1}^{n} H_n(\hat{F}_l, F_q) \tag{5.8}$$

Using the same methodology, the distance between each majority sample and the synthetic instance is determined. The minimal distance produced samples is then removed in order to get high purity data. These approaches indicate that the dataset contains no increased noisy or borderline samples.

6.2.3 Student Performance Prediction using Explainable Deep Learning based Knowledge Distillation Framework

The study introduces an innovative EDL_KDF framework for predicting student performance. While complex deep learning models can attain high accuracy, they often come with the drawback of increased computational complexity due to a big No. of attributes. In distinction, Knowledge Distillation (KD) offers a way to accurately forecast student achievement with lower computational overhead. KD achieves this by distilling knowledge from a larger, heavy-weight model into a more efficient, light-weight model, thereby reducing the model's complexity without compromising its accuracy. This approach allows for a more streamlined and efficient prediction process while maintaining strong performance metrics. The suggested framework structure is depicted in **Fig. 6.2**.

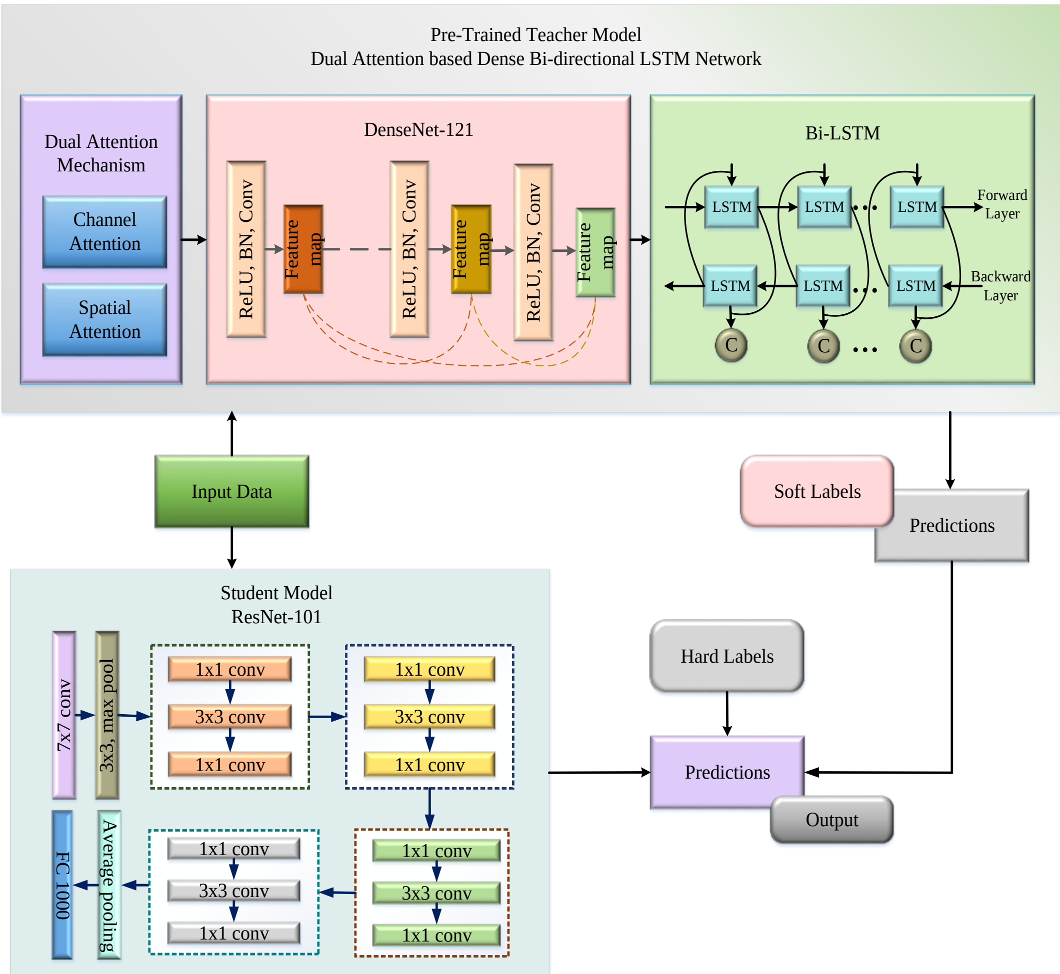


Fig. 6.2 Structure of EDL_KDF framework

By applying the condensed knowledge, KD used the pre-trained instructor model (big model) to assist in training the student model (small model). KD has two loss components for student model training. Decreases the difference among the expected and actual labels (hard labels) in the first place, but decreases the teacher's predictions in the second place (soft labels). While distillation process, the probability value is attained using the temperature (t). Initially the logits is extracted from the student model for calculate the probability then it feed in the softmax function along with t. Softened softmax function using the temperature is given as below.

$$Soft\max(v_i) = \frac{e(u_i / t)}{\sum_{k=1}^{l} e(u_k / t)} \tag{5.9}$$

The input data is initially given to the both student and teacher model and receive the logits v_z and v_t. Student and teacher model's probability distribution calculated using following formula.

$$p(t) = Soft\max(u_t / t) \tag{5.10}$$

$$p(Z) = Soft\max(v_z / t) \tag{5.11}$$

The probability of teacher and student entropy are control by using the t. The distillation loss can be expressed as below.

$$l = \lambda * lh(p(Z), O) + (1 - \lambda) * lz(p(Z), p(t)) \tag{5.12}$$

Where, O, λ, are denoted as actual label (hard label) and weight coefficient respectively. lh and lz are computed through the cross entropy. The student learns the appropriate probabilities of specific categories using the pretrained teacher model using an increasing entropy in $p(Z)$ accordingly with t. The teacher model and student models are mentioned in the below sub-section.

6.2.3.1 Pre-trained Teacher Network using Dual Attention based Dense Bi-directional LSTM Network

LSTM is a type of RNN architecture intended to model sequences and temporal data effectively. Unlike conventional RNNs, LSTMs are prepared with memory cells and a sophisticated gating mechanism that allows them to learn and retain long-term dependences in data while mitigating the disappearing gradient issue commonly encountered in DNNS.

DAbDBi-LSTM Network, used as the teaching model, comprises DenseNet-121, Bi-LSTM, and a Dual Attention Mechanism. This advanced network architecture leverages the strengths of each component to enhance prediction performance. DenseNet-121 serves as a robust convolutional neural network for feature extraction, while Bi-LSTM provides powerful sequence modeling capabilities. The Dual Attention Mechanism integrates spatial and channel attention, allowing the network to effort on both the spatial spreading of features and their importance across different channels. This combined attention approach enhances the network's ability to accurately capture and weigh relevant information, leading to improved teaching and learning performance in the technique.

The spatial data of feature maps is initially aggregated using average pooling with maximum pooling for channel attention that producing two separate spatial feature maps such as average pooling (D_{Ag}^{z})and max pooling (D_{Mx}^{z}).After that, an integrated network receives this feature map, producing the channel attention $G_S \in \Re^{z \times 1 \times 1}$. The network have hidden layer and multilayer perceptron. By using the reduction ratio the hidden activation size is decreased as $\Re^{z/r \times 1 \times 1}$ for minimize the parameter overhead. Element summation is employed to create a feature output vector for every feature map and channel attention is expressed as below.

$$G_z(D) = Sigmoid\left(MLP(D_{Ag}^{z}) + MLP(D_{Mx}^{z})\right) \tag{5.12}$$

The channel information of the feature map is combined using two pooling methods for spatial attention that producing two 2D feature maps such as average pooling ($D_{Ag}^{K} \in \Re^{1 \times W \times G}$) and maximum pooling $D_{Mx}^{K} \in \Re^{1 \times w \times G}$. Next, create spatial attention map by joining the two feature maps onto a convolutional layer that has been expressed as below.

$$G_g(D) = Sigmoid\left(f^{7 \times 7}\left[D_{Ag}^{z}; D_{Mx}^{z}\right]\right) \tag{5.13}$$

Where, $f^{7 \times 7}$ denoted the convolution kernel with 7×7 kernel operation. DenseNet-121 architecture contained the dense connection with 121 layers, which make connection among the all following layers. To improve the efficiency of the network's training, each layer obtains significant features that have been learned by previous layers. . It utilized less amount of parameter while training and it solve overfitting problem using dense connection. The dense block is the important part of the DenseNet that have Batch Normalization, ReLU and 3×3 Conv that can be expressed as below.

$$C_i = w_i([C_0, C_1, \ldots C_{i-1}]) \tag{5.14}$$

Where, layers are denoted as $(0,1,\ldots i-1)$ the feature map conjunction generated by DenseNet-121 layers are represented as $[C_0, C_1, \ldots C_{i-1}]$. Operation done over $i-th$ layer is denoted as $W_i(.)$ thus, the DenseNet effectively learn the feature map. To realize the deep features produced by the DenseNet-121, Bi-LSTM is introduced in the student model. The LSTM contained three gates such input (Z_g), forget (t_g) and output (S_g) gates. For timestep x, the LSTM expression are derived as below.

$$\tilde{C} = \tanh\left(K_g U_{yc} + aE_{g-1} U_{hc} + Bz_g\right) \tag{5.15}$$

Where, Bz_g denoted as bias parameter, U_{yc} and U_{hc} represented as weight parameters. In LSTM, S_g note that new data get using D_g. D_{g-1} contained old memory count is addressed by using T_g and by using the pointwise multiplication the equation updated as follows.

$$C_g = C_g \otimes C_{g-1} + Z_g \otimes \tilde{C}_g \tag{5.16}$$

Information is transferred between the memory cell and the concealed state via an output gate.

$$t_g = R_g \otimes \tanh(C_g) \tag{5.17}$$

The Bi-LSTM process based on the LSTM but it receive the input data in both direction. The Bi-LSTM at the timestep y is expressed as below.

$$\overrightarrow{t}_x = \phi\left(K_x U_{yx}^{(f)} + \overrightarrow{E}_{x-1} U_{hx}^{(f)} + Bz_x^{(f)} \right)$$
$$\overleftarrow{t}_x = \phi\left(K_x U_{yx}^{(b)} + E\overleftarrow{x}_{y-1} U_{hx}^{(b)} + Bz_x^{(b)} \right) \tag{5.18}$$

The hidden state is obtained by merging the forward and backward LSTM layers, after which it is sent to the output layer, which is stated as follows.

$$R_g = K_g U_{hq} + Bz_q \tag{5.19}$$

Where, Bz_q is a bias parameter and U_{hq} is a weight parameters are the output layer's model parameters. Finally the softmax function predicted the student performance. Thus the model effectively learned the student performance data's presented in the dataset through the DAbDBi-LSTM Network.

6.2.3.2 Prediction make using ResNet-101 Student Network

The ResNet-101 architecture consists of 101 layers and is dependent on the learning methodology of Residual neural networks. It is one of the deepest architectures proposed for ImageNet. What sets ResNet-101 apart from other architectures is its optimization of residuals between input and desired convolution properties. This approach enables the desired features to be obtained more efficiently and effectively compared to other architectures. By optimizing residuals, ResNet-101 can significantly lessen the No. of attributes in a deep network. This reduction in parameters not only streamlines the model but also allows the number of layers to be brought down to a more effective level. As a result, ResNet-101 can maintain a high level of performance while simultaneously improving computational efficiency.

ResNet-101 optimizes residuals across input and targeted convolution feature sets, utilizing 101 layers for efficient attainment of desired properties compared to other networks. Through residual optimization, ResNet-101 effectively manages the parameter count of the deep network, thereby reducing the network's complexity and the total No. of layers required. The architecture employs ResBlock layers, which transfer data from earlier layers of the network to the current layer. These ResBlock layers feed residual values from higher layers back to lower layers, facilitating smoother information flow and aiding in the training process. The two-layer activation in ResNet-101 consists of the ReLU activation function and a skip connection that passes value through various weight levels, impacting the system's overall performance. This skip connection, in combination with the nonlinear ReLU function, plays a critical role in the network's skill to handle deep structures effectively.

$$R = w_{2\sigma}(w_1 u) \tag{5.20}$$

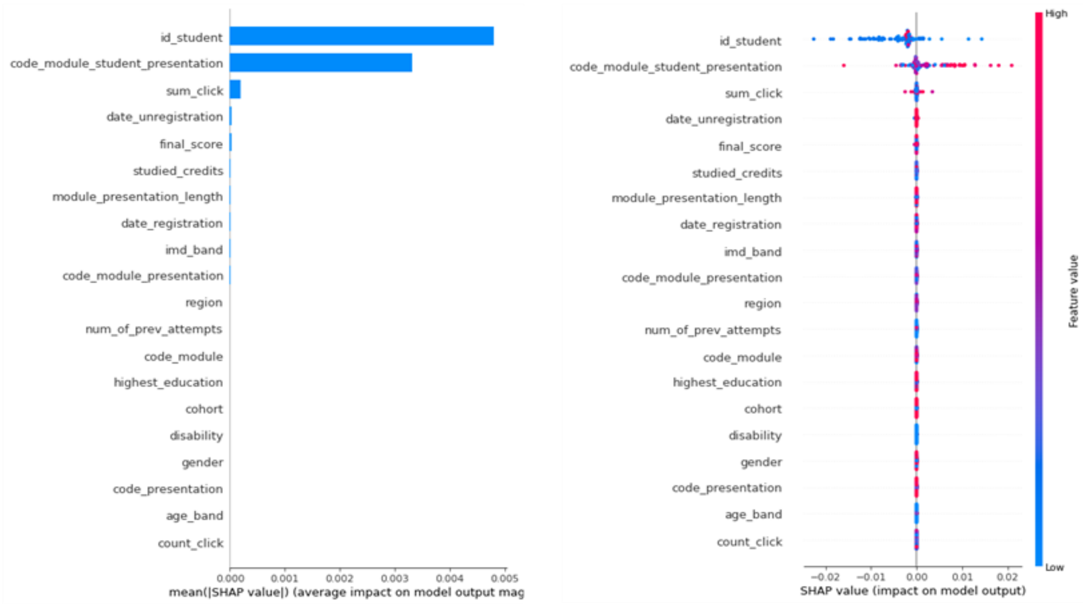
Output is attained by concatenate the second Relu value, the output can be expressed as below.

$$O = R(v, \{w_i\}) + u \tag{5.21}$$

Where, u and O are represented the input and output vector respectively. At the end the softmax activation function predicted the student performance by using the ResNet-101 with the help of DAbDBi-LSTM teacher network.

6.2.3.3 SHapley Additive Explanations

In order to examine the feature significant contribution for the prediction outcome, SHAP is typically employed. A heatmap shows how feature values affect the result of the prediction. The model's characteristics, mean absolute Shapley value, and magnitude outcome are displayed in **Fig. 6.3**.



(a) (b)

Fig. 6.3 Analysis of (a) Mean absolute and (b) magnitude outcome of SHAP value

The SHAP analysis demonstrates that the feature "id_student" has a feature magnitude around 0.005, indicating its predictive effect on student performance. This feature's low values, which vary between -0.02 and 0.014, show both positive and negative impacts on the proposed prediction model. In contrast, the "code module student presentation" feature's low values positively influence predictions, increasing them from -0.006 to 0.02. SHAP supports the understanding of how different features impact the EDL_KDF prediction model's output by highlighting the relevance and influence of each characteristic within the model. This analysis provides insight into the model's decision-making procedure and the relative importance of different features.

6.3 RESULT AND DISCUSSION

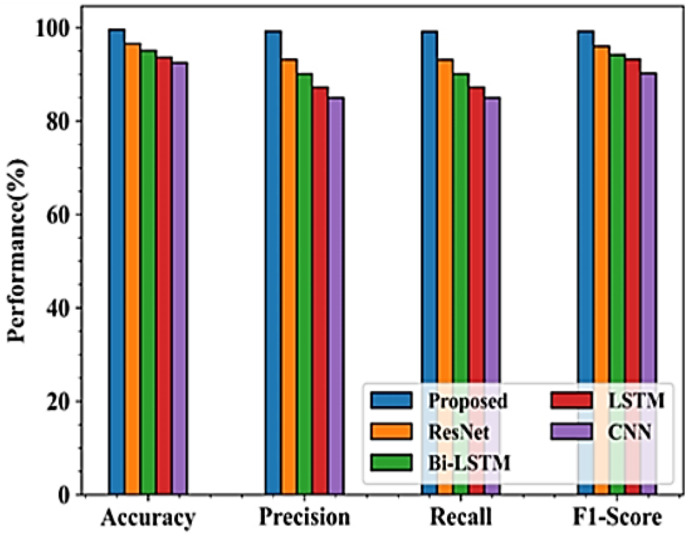
The efficacy of the suggested work is demonstrated in this part by evaluating the EDL_KDF framework's performance in comparison to other current student performance prediction frameworks. The Python programming language and its related libraries were utilized to create the suggested EDL_KDF framework, facilitating effective data management and model creation. This method allows for a comprehensive evaluation of the framework's capabilities and benefits over alternative approaches.

6.3.1 Dataset description

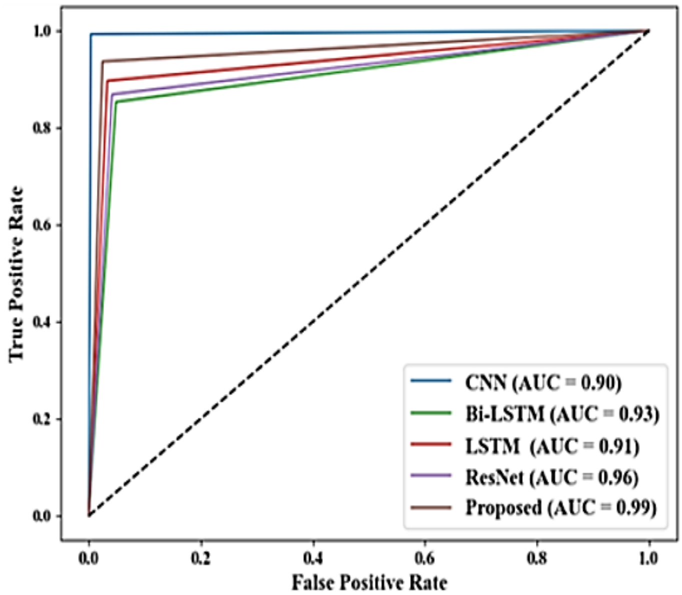
A standardization and benchmarking dataset is the OULAD. 32,593 student log data, including demographics, interaction data, registrations, and assignments, were included in the dataset. The dataset is download using https://www.kaggle.com/datasets/anlgrbz/student-demographics-online-education-dataoulad link. The Open University takes into account courses in the social sciences as well as STEM (science, technology, engineering, and mathematics) to compile data on student engagement. Data on student activities are available for the years 2013–2014.

6.3.2 Analysis of result and comparison

Several deep learning performance indicators, such as accuracy, precision, recall, F1-score, and ROC Curve, are utilized to assess the efficacy of the EDL_KDF architecture. These indicators offer a thorough evaluation of the framework's effectiveness. While precision assesses the percentage of real positive forecasts to all positive predictions, accuracy gauges how accurate predictions are overall. The model's recall measures how well it can recognize real positive examples; the F1-score combines recall and accuracy to give a fair assessment of the model's performance. The ROC Curve analyzes the trade-off among true positive and false positive rates across different threshold levels, offering insights into the model's ability to distinguish between classes. Together, these metrics offer a detailed understanding of the EDL_KDF framework's performance in predicting student outcomes. These analytic graphs are shown in **Fig. 6.4**.



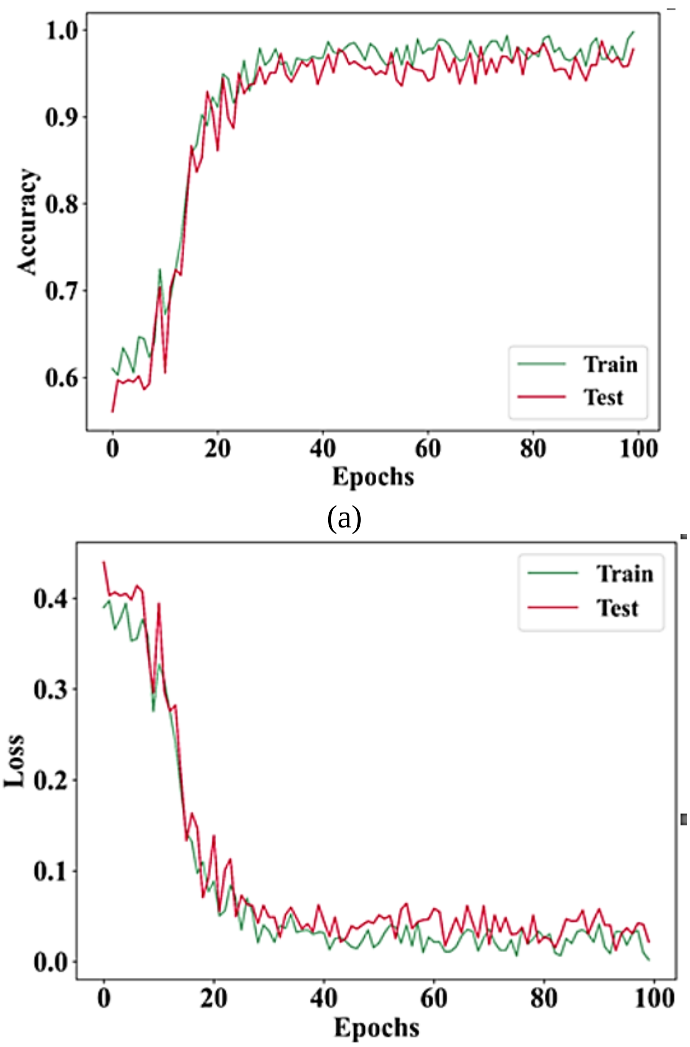
(a)



(b)

Fig. 6.4 Analysis of EDL_KDF framework using (a) various performance metrics (b) ROC curve

It is usual practice to evaluate the prediction replica's efficacy, resilience, and efficiency using the DL execution measures. The DL model outperformed in the prediction due to its strong F1-score, precision, recall, and accuracy outcomes. The EDL_KDF framework achieved an F1-score of 99.6%, accuracy, precision, and recall of 99.21%. The framework also achieved 0.99 Area under Curve (AUC) in ROC analysis at the same time. When compared to other DL models, the performance that was attained yielded better results. The accuracy and loss curve analysis are displayed in **Fig. 6.5**.



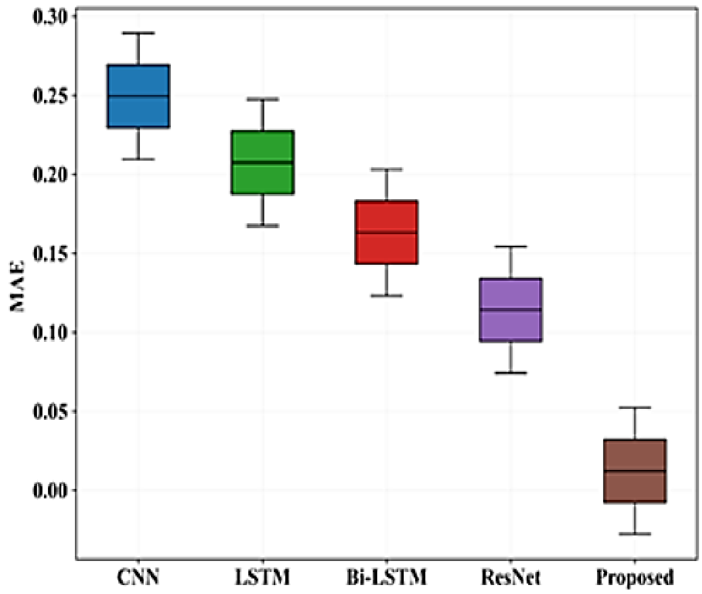
(a)

(b)

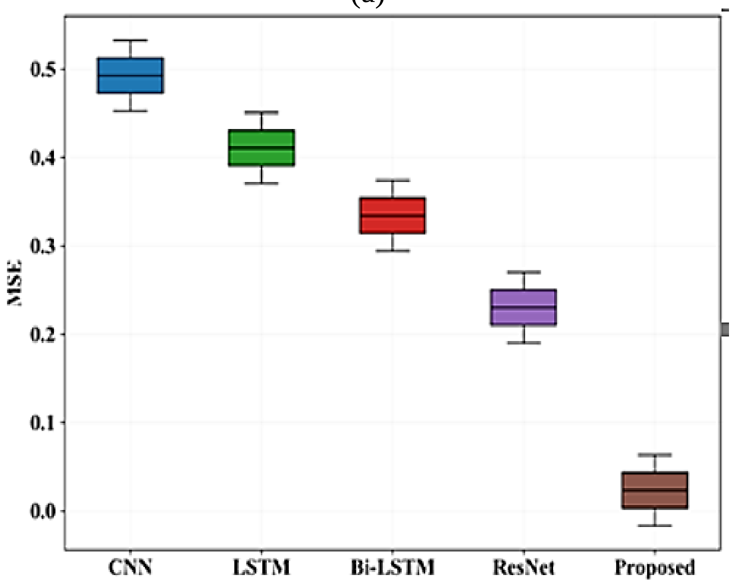
Fig. 6.5 Analysis of (a) Accuracy curve and (b) Loss curve

The performance of the EDL_KDF framework is evaluated over 100 epochs, focusing on accuracy and loss for both training as well as testing phases. After the first 40 epochs, the prediction model shows a notable decrease in loss, reflecting improved learning and model performance. Simultaneously, the model is effectively catching trends in the data, as seen by the forecast accuracy's steady rise.

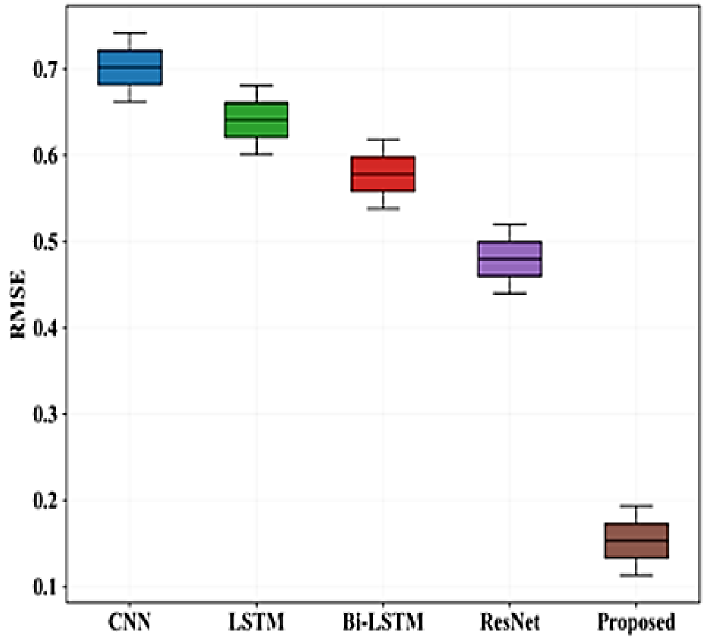
This progressive improvement in the model's accuracy as well as reduction in loss across the epochs supports its ability to accurately forecast student performance using the OULAD dataset. This highlights the framework's potential for successful application in predicting student outcomes and performance. The execution analysis of the recommended framework in rate of error metrics is displayed in **Fig. 6.6**.



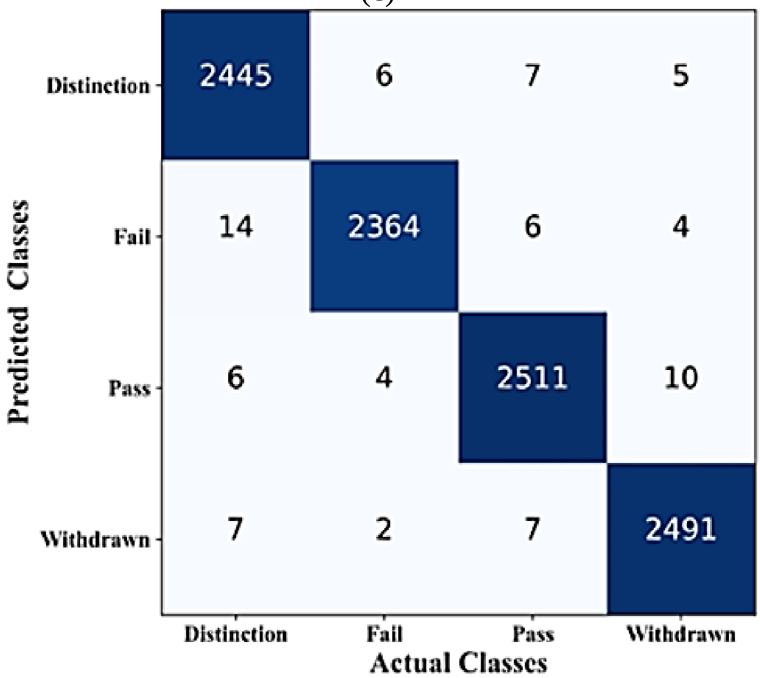
(a)



(b)



(c)



(d)

Fig. 6.6 Comparison and performance analysis based on (a) MAE (b) MSE (c) RMSE and (d) Confusion matrix

The error rate of the prediction model is typically assessed using metrics such as MSE,MAE, and RMSE. The EDL_KDF framework demonstrates superior performance, resulting in notably low values for these error metrics, including 0.023% for MSE, 0.0122% for MAE, and 0.1528% for RMSE. These low error values indicate the model's high precision in predicting student performance, with accurate classification of different student performance categories. Moreover, inaccurate predictions across the various classes are relatively rare, showcasing the model's overall excellence. Through this comprehensive analysis, the EDL_KDF framework stands out as highly effective in forecasting student performance.

6.3.3 Discussion

This section explores the performance of the EDL_KDF framework and its effectiveness in overcoming challenges associated with current student performance prediction models. The student model within the EDL_KDF framework incorporates DenseNet-121, which mitigates the issue of overfitting. Additionally, the use of ISMOTE resolves the class imbalance problem in the dataset. The novel EDL_KDF framework addresses concerns regarding prediction accuracy and computational complexity. By utilizing the knowledge distillation from the DAbDBi-LSTM network, the framework leverages the ResNet-101 network for prediction, effectively reducing computational complexity while maintaining high accuracy in predictions. This approach streamlines the process and enhances the overall performance of student performance forecasting. The comparative study with the current prediction model was shown in **Table 6.1**.

Table 6.1 Comparison analysis of accuracy with existing student prediction model

Author and References	Model	Accuracy (%)
Waheed et al. [111]	Deep ANN	84-93
Giannakas et al. [112]	DNN	82.39
Asselman et al. [113]	Adaboost, XGBoost and random forest ensembled algorithm	71.40
Rodríguez-Hernández et al. [114]	ANN	82
Kusumawardani et al. [115]	DL based transformer encoder	83.17
Kukkar et al. [116]	RNN + LSTM + RF	97
Almahdi et al. [117]	DL based an Optimized Predictive Academic Performance Approach	98.94
Proposed	**EDL_KDF framework**	**99.6**

Table 6.2 Comparison analysis of precision with existing student prediction model

Author and References	**Model**	**Precision (%)**
Waheed et al. [111]	Deep ANN	93-96
Giannakas et al. [112]	DNN	77.08
Asselman et al. [113]	Adaboost, XGBoost and random forest ensembled algorithm	73.97
Proposed	**EDL_KDF framework**	**99.21**

Table 6.3 Comparison analysis of recall with existing student prediction model

Author and References	**Model**	**Precision (%)**
Waheed et al. [111]	Deep ANN	69-86
Giannakas et al. [112]	DNN	68.52
Asselman et al. [113]	Adaboost, XGBoost and random forest ensembled algorithm	78.15
Proposed	**EDL_KDF framework**	**99.21**

The study showed that the proposed EDL_KDF framework excels in expecting student performance with high accuracy, precision, and recall, even with the complexity of the data and computation involved. This advanced approach proves to be more effective than other existing prediction models, as it maintains a high level of performance across multiple metrics. The EDL_KDF framework's ability to handle intricate patterns in the data while delivering reliable results sets it apart from other models, making it a superior choice for student performance forecasting.

6.4 SUMMARY

The suggested EDL_KDF framework in this study successfully forecasted the OULAD dataset's student performance. Predictive accuracy was originally enhanced by the EDL_KDF framework by attractive the value of the dataset during the pre-processing phase. Following that, the new EDL_KDF framework used the KD model to predict student performance with accuracy. With the assistance of the DAbDBi-LSTM network and ResNet-101, the EDL_KDF framework was able to achieve a higher level of accuracy. In order to analyze each feature in the model and determine how a given characteristic affects the prediction, SHAP was finally created. The EDL_KDF framework was assessed using a number of performance criteria, and across the OULAD dataset, it obtained accuracy and precision of 99.6% and 99.21%, respectively.

CHAPTER 7

CONCLUSION, FUTURE SCOPE AND SOCIAL IMPACT

7.1 CONCLUSION

In the first objective, various ML approaches are used to apply TSCNDE work, and student performance is assessed and contrasted with recently created methodologies. Students' academic challenges can be resolved by applying the TSCNDE approach. The extremely accurate features utilised in the prediction procedure yield superior outcomes when measuring student achievement compared to the current approaches. The comparison outcome demonstrates suggested TSCNDE method's effectiveness. When predicting student performance and evaluating the PYTHON simulation tool, the TSCNDE approach yields results with 99.2% accuracy. The recommended TSCNDE method's corresponding precision rating is 99.8%, indicating that it can provide accurate findings for predicting student performance. In addition, the TSCNDE technique predicts student performance with a 98.9% sensitivity and a 98.7% specificity.

The second effort that proposes hybrid optimization (Gannet Hunt) combines the collaborative behavior of the bay-winged predator with the capturability of the gannet to capture prey in order to quickly arrive at the optimal global solution. To reduce the computational overhead in learning, the suggested Gannet Hunt method is first used to identify the best features. Additionally, by altering the adjustable parameters throughout the learning process, the Gannet Hunt algorithm's introduction lowers information loss and improves the LSTM's capacity for generalization. The maximum values of 99.9%, 99.5738%, 99.99%, and 99.78% were achieved by analysis based on metrics such as F-Measure, Recall, Precision, and Accuracy, respectively.

In the third work, data from prior students is initially subjected to an outlier identification technique. In big datasets, techniques based on distance and density are commonly used to identify outliers that are sensitive to k-values and lead to local density issues. A hybrid deep learning approach is used to obtain better features from entropy-based characteristics in order to produce dependable prediction. In the results section, the suggested model and the existing models are compared using the following metrics: accuracy, precision, recall, f1-score, MSE, and RMSE. The proposed model had the greatest accuracy range of 98.5% when related to other systems.

Recommended EDL_KDF framework in the final approach successfully forecasted the student performance across the OULAD dataset. Predictive accuracy was originally enhanced by the EDL_KDF framework by improving quality of the dataset during pre-processing phase. Following that, the new EDL_KDF framework used the KD unit to predict student performance with accuracy. With the assistance of the DAbDBi-LSTM network and ResNet-101, the EDL_KDF framework was able to achieve a higher level of accuracy. The EDL_KDF framework was assessed using a number of performance criteria, and across the OULAD dataset, it obtained accuracy and precision of 99.6% and 99.21%, respectively. In the future, the KD model's generalization loss will be concentrated on enhancing student performance prediction accuracy for unidentified data.

7.2 FUTURE SCOPE

- Expand the research by incorporating multiple datasets from different educational institutions to enhance the model's robustness and generalizability.
- Develop real-time monitoring and intervention strategies based on predictive insights to proactively support students' learning journeys.
- Investigate cross-disciplinary applications of the model, such as its potential use in mental health and well-being monitoring for students.
- Utilize a wider variety of datasets from different educational contexts to improve the generalizability of the model and its predictive capabilities.
- Explore the use of more advanced DL structures, such as graph neural networks, to improve model performance.
- Investigate application of the EDL_KDF framework in employee training and development to optimize performance in organizational settings.
- Continuously refine the SHAP analysis to improve interpretability and trustworthiness of the model's predictions.
- Develop adaptive learning systems that use predictions from the EDL_KDF framework to tailor interventions and support for students.

7.3 SOCIAL IMPACT

The social impact of the research work presented in this thesis is immense in the domain of student performance prediction and optimization. Proposed approaches like the TSCNDE method, Hybrid Gannet Hunt optimization, and EDL_KDF framework showed high accuracy in predicting student performance to the tune of 98.5% to 99.9% across different metrics. The proposed models will be helpful in the early identification of at-risk students and will make timely interventions, and personalize support for improved academic outcomes. High accuracy predictions might raise decision-making for policymakers, administrators, and educators toward more equitable, inclusive educational practices. Accurate predictions of student performance allow educators to optimally allocate resources and interventions, ensuring students from all backgrounds have the opportunity to succeed. The optimization techniques allow for abstracting the feature engineering and model selection steps, hence making these proposed models usable for a larger user base. The accuracy of such early predictions of student performance allows for timely interventions, in which educators are able to offer tailored support opportunities and learning for appropriate students. It would connote that students will be more engaged, there would be less dropout, and students would function at an overall higher academic achievement, which would be beneficial to the school. By this research work, the proposed models can also help educators in discovering and tackling challenges faced by unique populations within diverse students, which will definitely improve more equitable and inclusive educational practices. Such high-accuracy prediction models may benefit associated stakeholders, such as administrators, policymakers, and educators, by providing them with insights into the execution of data-driven decisions that will promote the well-being of students.

REFERENCES

[1] Dlouhá, Jana, and Marie Pospíšilová. "Education for Sustainable Development Goals in Public Debate: The Importance of Participatory Research in Reflecting and Supporting the Consultation Process in Developing a Vision for Czech Education." *Journal of Cleaner Production*, vol. 172, Jan. 2018, pp. 4314–4327, https://doi.org/10.1016/j.jclepro.2017.06.145.

[2] Pulimeno, Manuela, et al. "School as Ideal Setting to Promote Health and Wellbeing among Young People." *Health Promotion Perspectives*, vol. 10, no. 4, 7 Nov. 2020, pp. 316–324, www.ncbi.nlm.nih.gov/pmc/articles/PMC7723000/, https://doi.org/10.34172/hpp.2020.50.

[3] Larrabee Sønderlund, Anders, et al. "The Efficacy of Learning Analytics Interventions in Higher Education: A Systematic Review." *British Journal of Educational Technology*, vol. 50, no. 5, 14 Nov. 2018, pp. 2594–2618, https://doi.org/10.1111/bjet.12720.

[4] Alyahyan, Eyman, and Dilek Düştegör. "Predicting Academic Success in Higher Education: Literature Review and Best Practices." *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, 10 Feb. 2020, educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-020-0177-7, https://doi.org/10.1186/s41239-020-0177-7.

[5] Rastrollo-Guerrero, Juan L., et al. "Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review." *Applied Sciences*, vol. 10, no. 3, 4 Feb. 2020, p. 1042, https://doi.org/10.3390/app10031042.

[6] Hamoud, Alaa Khalaf, et al. "Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis." *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 2, 2018, p. 26, https://doi.org/10.9781/ijimai.2018.02.004. Accessed 12 Dec. 2019.

[7] Bunce, Louise, et al. "The Student-As-Consumer Approach in Higher Education and Its Effects on Academic Performance." *Studies in Higher Education*, vol. 42, no. 11, 14 Jan. 2017, pp. 1958–1978, www.tandfonline.com/doi/pdf/10.1080/03075079.2015.1127908, https://doi.org/10.1080/03075079.2015.1127908.

[8] Kavi, Avinash, and Padmaja R. Walvekar. "Lifestyle Factors Influencing the Academic Performance among the Secondary School Students in an Urban Area of South India." *International Journal of Adolescent Medicine and Health*, vol. 0, no. 0, 7 Sept. 2020, https://doi.org/10.1515/ijamh-2020-0091.

[9] Greenhow, Christine, and Cathy Lewin. "Social Media and Education: Reconceptualizing the Boundaries of Formal and Informal Learning." *Learning, Media and Technology*, vol. 41, no. 1, 16 July 2015, pp. 6–30, https://doi.org/10.1080/17439884.2015.1064954.

[10] Hosen, Mosharrof, et al. "Individual Motivation and Social Media Influence on Student Knowledge Sharing and Learning Performance: Evidence from an Emerging Economy." *Computers & Education*, vol. 172, Oct. 2021, p. 104262, https://doi.org/10.1016/j.compedu.2021.104262.

[11] Khasanah, Yunita Uswatun. "A Pragmatic Analysis of Efl Learners' Social Media Interaction." *JETLe (Journal of English Language Teaching and Learning)*, vol. 1, no. 1, 31 Oct. 2019, p. 27, https://doi.org/10.18860/jetle.v1i1.7751.

[12] AlFaris, Eiad, et al. "The Pattern of Social Media Use and Its Association with Academic Performance among Medical Students." *Medical Teacher*, vol. 40, no. sup1, 6 May 2018, pp. S77–S82, https://doi.org/10.1080/0142159x.2018.1465536.

[13] Kucharska, Wioleta. "Consumer Social Network Brand Identification and Personal Branding. How Do Social Network Users Choose among Brand Sites?" *SSRN Electronic Journal* , vol. 4, no. 1, 2017, https://doi.org/10.2139/ssrn.2954929.

[14] Fung, XavierC. C., et al. "The Effects of Internet Gaming and Social Media Use on Physical Activity, Sleep, Quality of Life, and Academic Performance among University Students in Hong Kong: A Preliminary Study." *Asian Journal of Social Health and Behavior*, vol. 4, no. 1, 2021, p. 36, https://doi.org/10.4103/shb.shb_81_20. Accessed 16 Mar. 2021.

[15] Holtzblatt, Lester, and Mary Lou Tierney. "Measuring the Effectiveness of Social Media on an Innovation Process." *Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '11*, 2011, https://doi.org/10.1145/1979742.1979669. Accessed 6 Dec. 2022.

[16] Dennen, Vanessa P., and Kerry J. Burner. "Identity, Context Collapse, and Facebook Use in Higher Education: Putting Presence and Privacy at Odds." *Distance Education*, vol. 38, no. 2, 4 May 2017, pp. 173–192, https://doi.org/10.1080/01587919.2017.1322453. Accessed 21 May 2020.

[17] Ilin, Vladislav. "The Role of User Preferences in Engagement with Online Learning." *E-Learning and Digital Media*, vol. 19, no. 2, 3 Aug. 2021, p. 204275302110355, https://doi.org/10.1177/20427530211035514.

[18] Chen, Bodong, et al. "Fostering Student Engagement in Online Discussion through Social Learning Analytics." *The Internet and Higher Education*, vol. 37, Apr. 2018, pp. 21–30, https://doi.org/10.1016/j.iheduc.2017.12.002.

[19] McCarthy, Claudine. "Digital Storytelling on Social Media Brings Assessment Data to Life." *Student Affairs Today*, vol. 24, no. 10, 19 Dec. 2021, pp. 1–6, https://doi.org/10.1002/say.31011.

[20] Paakkari, Leena, et al. "Problematic Social Media Use and Health among Adolescents." *International Journal of Environmental Research and Public Health*, vol. 18, no. 4, 15 Feb. 2021, pubmed.ncbi.nlm.nih.gov/33672074/, https://doi.org/10.3390/ijerph18041885.

[21] A. Çetinkaya and Ömer Kaan Baykan, "Prediction of middle school students' programming talent using artificial neural networks," *Engineering Science and Technology, an International Journal*, vol. 23, no. 6, pp. 1301–1307, Dec. 2020, doi: https://doi.org/10.1016/j.jestch.2020.07.005.

[22] M. Mohammadi, M. Dawodi, W. Tomohisa, and N. Ahmadi, "Comparative study of supervised learning algorithms for student performance prediction," *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Feb. 2019, doi: https://doi.org/10.1109/icaiic.2019.8669085.

[23] Rahul and R. Katarya, "A Review: Predicting the Performance of Students Using Machine learning Classification Techniques," *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Dec. 2019, doi: https://doi.org/10.1109/i-smac47947.2019.9032493.

[24] A. Rajak, A. K. Shrivastava, and Vidushi, "Applying and comparing machine learning classification algorithms for predicting the results of students," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 23, no. 2, pp. 419–427, Feb. 2020, doi: https://doi.org/10.1080/09720529.2020.1728895.

[25] E. Tanuar, Y. Heryadi, Lukas, B. S. Abbas, and F. L. Gaol, "Using Machine Learning Techniques to Earlier Predict Student's Performance," *IEEE Xplore*, Sep. 01, 2018. https://ieeexplore.ieee.org/document/8626856 (accessed Dec. 09, 2020).

[26] A. Jain and S. Solanki, "An Efficient Approach for Multiclass Student Performance Prediction based upon Machine Learning," *2019 International Conference on Communication and Electronics Systems (ICCES)*, Jul. 2019, doi: https://doi.org/10.1109/icces45898.2019.9002038.

[27] A. Tripathi, S. Yadav, and R. Rajan, "Naive Bayes Classification Model for the Student Performance Prediction," *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Jul. 2019, doi: https://doi.org/10.1109/icicict46008.2019.8993237.

[28] H. A. Mengash, "Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems," *IEEE Access*, vol. 8, pp. 55462–55470, 2020, doi: https://doi.org/10.1109/access.2020.2981905.

[29] K. K. and J. S. Raj, "BIG DATA ANALYTICS FOR DEVELOPING SECURE INTERNET OF EVERYTHING," *Journal of ISMAC*, vol. 01, no. 02, pp. 49–56, Sep. 2019, doi: https://doi.org/10.36548/jismac.2019.2.006.

[30] S. Amjad, M. Younas, M. Anwar, Q. Shaheen, M. Shiraz, and A. Gani, "Data Mining Techniques to Analyze the Impact of Social Media on Academic Performance of High School Students," *Wireless Communications and Mobile Computing*, vol. 2022, p. e9299115, Mar. 2022, doi: https://doi.org/10.1155/2022/9299115.

[31] I. K. Nti, S. Akyeramfo-Sam, B. Bediako-Kyeremeh, and S. Agyemang, "Prediction of social media effects on students' academic performance using Machine Learning Algorithms (MLAs)," *Journal of Computers in Education*, Aug. 2021, doi: https://doi.org/10.1007/s40692-021-00201-z.

[32] S. Poudyal, M. J. Mohammadi-Aragh, and J. E. Ball, "Prediction of Student Academic Performance Using a Hybrid 2D CNN Model," *Electronics*, vol. 11, no. 7, p. 1005, Mar. 2022, doi: https://doi.org/10.3390/electronics11071005.

[33] G. B. Brahim, "Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features," *Arabian Journal for Science and Engineering*, Jan. 2022, doi: https://doi.org/10.1007/s13369-021-06548-w.

[34] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm," *Interactive Learning Environments*, pp. 1–20, May 2021, doi: https://doi.org/10.1080/10494820.2021.1928235.

[35] S. S. Kusumawardani and S. A. I. Alfarozi, "Transformer Encoder Model for Sequential Prediction of Student Performance Based on Their Log Activities," *IEEE Access*, vol. 11, pp. 18960–18971, 2023, doi: https://doi.org/10.1109/access.2023.3246122.

[36] A. Rivas, A. González-Briones, G. Hernández, J. Prieto, and P. Chamoso, "Artificial neural network analysis of the academic performance of students in virtual learning environments," *Neurocomputing*, May 2020, doi: https://doi.org/10.1016/j.neucom.2020.02.125.

[37] C. F. Rodríguez-Hernández, M. Musso, E. Kyndt, and E. Cascallar, "Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100018, Jan. 2021, doi: https://doi.org/10.1016/j.caeai.2021.100018.

[38] K. Shahzadi, M. S. Sarfraz, M. Saqib, and H. M. Shah, "Machine Learning Methods for Analysing the Impact of Social Media on Students Academic Performance," *Journal of Applied Artificial Intelligence*, vol. 1, no. 1, pp. 13–22, Dec. 2020, doi: https://doi.org/10.48185/jaai.v1i1.96.

[39] Yazn Alshamaila, M. Habib, Ibrahim Aljarah, Hamad Alsawalqah, H. Faris, and Anas Ratib Alsoud, "An Intelligent Approach for the Effect of Social Media on Undergraduate Students Performance," Apr. 2020, doi: https://doi.org/10.1145/3397125.3397146.

[40] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Computers in Human Behavior*, vol. 104, p. 106189, Mar. 2020, doi: https://doi.org/10.1016/j.chb.2019.106189.

[41] F. Giannakas, C. Troussas, I. Voyiatzis, and C. Sgouropoulou, "A deep learning classification framework for early prediction of team-based academic performance," *Applied Soft Computing*, vol. 106, p. 107355, Jul. 2021, doi: https://doi.org/10.1016/j.asoc.2021.107355.

[42] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Transfer Learning from Deep Neural Networks for Predicting Student Performance," *Applied Sciences*, vol. 10, no. 6, p. 2145, Mar. 2020, doi: https://doi.org/10.3390/app10062145.

[43] C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," *Computers & Education*, vol. 131, pp. 22–32, Apr. 2019, doi: https://doi.org/10.1016/j.compedu.2018.12.006.

[44] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using machine learning to predict student difficulties from learning session data," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 381–407, Feb. 2018, doi: https://doi.org/10.1007/s10462-018-9620-8.

[45] J. Y. Chung and S. Lee, "Dropout early warning systems for high school students using machine learning," *Children and Youth Services Review*, vol. 96, pp. 346–353, Jan. 2019, doi: https://doi.org/10.1016/j.childyouth.2018.11.030.

[46] R. Ghorbani and R. Ghousi, "Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques," *IEEE Access*, vol. 8, pp. 67899–67911, 2020, doi: https://doi.org/10.1109/access.2020.2986809.

[47] H. Pallathadka, A. Wenda, E. Ramirez-Asís, M. Asís-López, J. Flores-Albornoz, and K. Phasinam, "Classification and prediction of student performance data using various machine learning algorithms," *Materials Today: Proceedings*, Jul. 2021, doi: https://doi.org/10.1016/j.matpr.2021.07.382.

[48] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, Mar. 2022, doi: https://doi.org/10.1186/s40561-022-00192-z.

[49] A. Namoun and A. Alshanqiti, "Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review," *Applied Sciences*, vol. 11, no. 1, p. 237, Dec. 2020, doi: https://doi.org/10.3390/app11010237.

[50] P. Chakraborty, Mohammad Abu Yousuf, and S. Rahman, "Predicting Level of Visual Focus of Human's Attention Using Machine Learning Approaches," *Advances in intelligent systems and computing*, pp. 683–694, Dec. 2020, doi: https://doi.org/10.1007/978-981-33-4673-4_56.

[51] U. Pujianto, W. Agung Prasetyo, and A. Rakhmat Taufani, "Students Academic Performance Prediction with k-Nearest Neighbor and C4.5 on SMOTE-balanced data," *IEEE Xplore*, Dec. 01, 2020. https://ieeexplore.ieee.org/abstract/document/9315439 (accessed May 10, 2023).

[52] K. Shaheed *et al.*, "DS-CNN: A pre-trained Xception model based on depth-wise separable convolutional neural network for finger vein recognition," *Expert Systems with Applications*, vol. 191, p. 116288, Apr. 2022, doi: https://doi.org/10.1016/j.eswa.2021.116288.

[53] M. H. Qais, H. M. Hasanien, and S. Alghuwainem, "Transient search optimization: a new meta-heuristic optimization algorithm," *Applied Intelligence*, vol. 50, no. 11, pp. 3926–3941, Jul. 2020, doi: https://doi.org/10.1007/s10489-020-01727-y.

[54] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open University Learning Analytics dataset," *Scientific Data*, vol. 4, p. 170171, Nov. 2017, doi: https://doi.org/10.1038/sdata.2017.171.

[55] S. Hassan, H. Waheed, N. R. Aljohani, M. Ali, S. Ventura, and F. Herrera, "Virtual learning environment to predict withdrawal by leveraging deep learning," *International Journal of Intelligent Systems*, vol. 34, no. 8, pp. 1935–1952, May 2019, doi: https://doi.org/10.1002/int.22129.

[56] H. Heuer and A. Breiter, "Student Success Prediction and the Trade-Off between Big Data and Data Minimization," *DeLFI*, pp. 219–230, Jan. 2018.

[57] M. Hlosta, Z. Zdrahal, and J. Zendulka, "Ouroboros," *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, Mar. 2017, doi: https://doi.org/10.1145/3027385.3027449.

[58] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, "Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–21, Oct. 2018, doi: https://doi.org/10.1155/2018/6347186.

[59] S. Rizvi, B. Rienties, and S. A. Khoja, "The role of demographics in online learning; A decision tree based approach," *Computers & Education*, vol. 137, pp. 32–47, Aug. 2019, doi: https://doi.org/10.1016/j.compedu.2019.04.001.

[60] Astatke M, Weng C, Chen S (2023) A literature review of the effects of social networking sites on secondary school students' academic achievement. Interactive Learning Environments 31(4):2153-69.

[61] Setiawan R, Nath K, Cavaliere LP, Villalba-Condori KO, Arias-Chavez D, Koti K, Bagale G, Rajest SS (2021) The impact of teaching innovative strategy on academic performance in high schools (Doctoral dissertation, Petra Christian University).

[62] Brew EA, Nketiah B, Koranteng R (2021) A literature review of academic performance, an insight into factors and their influences on academic outcomes of students at senior high schools. Open Access Library Journal 8(6):1-4.

[63] Aslam N, Khan I, Alamri L, Almuslim R (2021) An improved early student's academic performance prediction using deep learning. International Journal of Emerging Technologies in Learning (iJET) 16(12):108-22.

[64] Khan A, Ghosh SK (2021) Student performance analysis and prediction in classroom learning: A review of educational data mining studies. Education and information technologies 26:205-40.

[65] Prabu M, Srivastava K, Sharma VK, Prabakar DD (2022) Evaluation and semantic approach for student performance prediction using data mining techniques. InAIP Conference Proceedings 2405(1). AIP Publishing.

[66] Hussain S, Khan MQ (2023) Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. Annals of data science 10(3):637-55.

[67] Ahmad S, El-Affendi MA, Anwar MS, Iqbal R (2022) Potential future directions in optimization of students' performance prediction system. Computational Intelligence and Neuroscience 2022.

[68] Tjandra E, Kusumawardani SS, Ferdiana R (2022) Student performance prediction in higher education: A comprehensive review. In AIP Conference Proceedings 2470(1):050005. AIP Publishing LLC.

[69] Okewu E, Adewole P, Misra S, Maskeliunas R, Damasevicius R (2021) Artificial neural networks for educational data mining in higher education: A systematic literature review. Applied Artificial Intelligence 35(13):983-1021.

[70] Alsharhan A, Salloum SA, Aburayya A (2021) Using e-learning factors to predict student performance in the practice of precision education. Pt. 2 J. Legal Ethical & Regul. Isses 24:1.

[71] Albreiki B, Zaki N, Alashwal H (2021) A systematic literature review of student'performance prediction using machine learning techniques. Education Sciences 11(9):552.

[72] Nurudeen M, Abdul-Samad S, Owusu-Oware E, Tanye HA (2022) Measuring the effect of social media on student academic performance using a social media influence factor model. Education and Information Technologies 1-24.

[73] Katarya R (2022) Deep auto encoder based on a transient search capsule network for student performance prediction. Multimedia Tools and Applications 1-25.

[74] Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H (2019) Harris hawks optimization: Algorithm and applications. Future generation computer systems 97:849-72.

[75] Shahzadi K, Sarfraz MS, Saqib M, Shah HM (2020) Machine Learning Methods for Analysing the Impact of Social Media on Students Academic Performance. Journal of Applied Artificial Intelligence 1(1):13-22.

[76] Altabrawee H, Ali OA, Ajmi SQ (2019) Predicting students' performance using machine learning techniques. JOURNAL OF UNIVERSITY OF BABYLON for pure and applied sciences 27(1):194-205.

[77] Xiong SY, Gasim EF, Ying CX, Wah KK, Ha LM (2022) A Proposed Hybrid CNN-RNN Architecture for Student Performance Prediction. International Journal of Intelligent Systems and Applications in Engineering 10(3):347-55.

[78] students' performance using machine learning techniques. IEEE Access 8, 67899-67911(2020).

[79] Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A.: Analyzing and predicting students' performance by means of machine learning: A review. Applied sciences 10(3), 1042(2020).

[80] Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K.: Classification and prediction of student performance data using various machine learning algorithms. Materials today: proceedings 80, 3782-3785(2023).

[81] Hussain, S., & Khan, M. Q.: Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. Annals of data science 10(3), 637-655(2023).

[82] Asselman, A., Khaldi, M., & Aammou, S.: Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments 31(6), 3360-3379(2023).

[83] Liu, Q., Huang, Z., Yin, Y., Chen, E., Xiong, H., Su, Y., & Hu, G.: Ekt: Exercise-aware knowledge tracing for student performance prediction. IEEE Transactions on Knowledge and Data Engineering 33(1), 100-115(2019).

[84] Lau, E. T., Sun, L., & Yang, Q.: Modelling, prediction and classification of student academic performance using artificial neural networks. SN Applied Sciences 1, 1-10(2019).

[85] Sultana, J., Rani, M. U., & Farquad, M. A. H.: Student's performance prediction using deep learning and data mining methods. Int. J. Recent Technol. Eng 8(1), 1018-1021(2019).

[86] Qiu, F., Zhang, G., Sheng, X., Jiang, L., Zhu, L., Xiang, Q., ... & Chen, P. K.: Predicting students' performance in e-learning using learning process and behaviour data. Scientific Reports 12(1), 453(2022).

[87] Rahul, & Katarya, R.: Deep auto encoder based on a transient search capsule network for student performance prediction. Multimedia Tools and Applications 82(15), 23427-23451(2023).

[88] https://www.kaggle.com/datasets/alchzamb/mendeley/discussion

[89] https://www.kaggle.com/datasets/anlgrbz/student-demographics-online-education-dataoulad

[90] Brahim, G. B.: Predicting student performance from online engagement activities using novel statistical features. Arabian Journal for Science and Engineering 47(8), 10225-10243(2022).

[91] Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O.: Transfer learning from deep neural networks for predicting student performance. Applied Sciences 10(6), 2145(2020).

[92] Rivas, A., Gonzalez-Briones, A., Hernandez, G., Prieto, J., & Chamoso, P.: Artificial neural network analysis of the academic performance of students in virtual learning environments. Neurocomputing 423, 713-720(2021).

[93] Poudyal, S., Mohammadi-Aragh, M. J., & Ball, J. E.: Prediction of student academic performance using a hybrid 2D CNN model. Electronics 11(7), 1005(2022).

[94] Wu, Z., Huang, L., Huang, Q., Huang, C., & Tang, Y.: SGKT: Session graph-based knowledge tracing for student performance prediction. Expert Systems with Applications 206, 117681(2022).

[96] Escueta, Maya, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos. "Education technology: An evidence-based review." (2017).

[97] Ulanday, Ma Leah, Zarah Jane Centeno, Ma Cristina Bayla, and Joseph Callanta. "Flexible learning adaptabilities in the new normal: E-learning resources, digital meeting platforms, online learning systems and learning engagement." *Asian Journal of Distance Education* 16, no. 2 (2021).

[98] Stecuła, Kinga, and Radosław Wolniak. "Influence of COVID-19 pandemic on dissemination of innovative e-learning tools in higher education in Poland." *Journal of Open Innovation: Technology, Market, and Complexity* 8, no. 2 (2022): 89.

[99] Liu, Zi-Yu, Natalya Lomovtseva, and Elena Korobeynikova. "Online learning platforms: Reconstructing modern higher education." *International Journal of Emerging Technologies in Learning (iJET)* 15, no. 13 (2020): 4-21.

[100] Dickinson, Amber. "Communicating with the online student: the impact of e-mail tone on student performance and teacher evaluations." *Journal of Educators Online* 14, no. 2 (2017): n2.

[101] Zeichner, Kenneth M. "Competition, Economic Rationalization, Increased Surveillance, and Attacks on Diversity: Neo-liberalism and the Transformation of Teacher Education in the US 1." In *The struggle for the soul of teacher education*, pp. 40-62. Routledge, 2017.

[102] Ghosh, Ratna, and Xiaoli Jing. "Fostering global citizenship through student mobility: COVID-19 and the 4th wave in internationalization of education." *Beijing International Review of Education* 2, no. 4 (2020): 553-570.

[103] Miotto, Giorgia, Cristina Del-Castillo-Feito, and Alicia Blanco-González. "Reputation and legitimacy: Key factors for Higher Education Institutions' sustained competitive advantage." *Journal of Business Research* 112 (2020): 342-353.

[104] Maxwell, Sophie, Katherine J. Reynolds, Eunro Lee, Emina Subasic, and David Bromhead. "The impact of school climate and school identification on academic achievement: Multilevel modeling with student and teacher data." *Frontiers in psychology* 8 (2017): 2069.

[105] Kuleto, Valentin, Milena Ilić, Mihail Dumangiu, Marko Ranković, Oliva MD Martins, Dan Păun, and Larisa Mihoreanu. "Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions." *Sustainability* 13, no. 18 (2021): 10424.

[106] Gray, Geraldine, Ana Elena Schalk, Gordon Cooke, Phelim Murnion, Pauline Rooney, and K. C. O'Rourke. "Stakeholders' insights on learning analytics: Perspectives of students and staff." *Computers & Education* 187 (2022): 104550.

[107] Aslan, Muhammet Fatih, Kadir Sabanci, Akif Durdu, and Muhammed Fahri Unlersen. "COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization." *Computers in biology and medicine* 142 (2022): 105244.

[108] Burman, Iti, and Subhranil Som. "Predicting students academic performance using support vector machine." In *2019 Amity international conference on artificial intelligence (AICAI)*, pp. 756-759. IEEE, 2019.

[109] Garg, Sachin, Abdul Aleem, and Manoj Madhava Gore. "Employing Deep Neural Network for Early Prediction of Students' Performance." In *Intelligent Systems: Proceedings of ICMIB 2020*, pp. 497-507. Springer Singapore, 2021.

[110] Rodríguez-Hernández, Carlos Felipe, Mariel Musso, Eva Kyndt, and Eduardo Cascallar. "Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation." *Computers and Education: Artificial Intelligence* 2 (2021): 100018.

[111] Waheed, Hajra, Saeed-Ul Hassan, Naif Radi Aljohani, Julie Hardman, Salem Alelyani, and Raheel Nawaz. "Predicting academic performance of students from VLE big data using deep learning models." *Computers in Human behavior* 104 (2020): 106189.

[112] Giannakas, Filippos, Christos Troussas, Ioannis Voyiatzis, and Cleo Sgouropoulou. "A deep learning classification framework for early prediction of team-based academic performance." *Applied Soft Computing* 106 (2021): 107355.

[113] Asselman, Amal, Mohamed Khaldi, and Souhaib Aammou. "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm." *Interactive Learning Environments* 31, no. 6 (2023): 3360-3379.

[114] Rodríguez-Hernández, Carlos Felipe, Mariel Musso, Eva Kyndt, and Eduardo Cascallar. "Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation." *Computers and Education: Artificial Intelligence* 2 (2021): 100018.

[115] Kusumawardani, Sri Suning, and Syukron Abu Ishaq Alfarozi. "Transformer Encoder Model for Sequential Prediction of Student Performance Based on Their Log Activities." *IEEE Access* 11 (2023): 18960-18971.

[116] Kukkar, Ashima, Rajni Mohana, Aman Sharma, and Anand Nayyar. "A novel methodology using RNN+ LSTM+ ML for predicting student's academic performance." *Education and Information Technologies* (2024): 1-37.

[117] Almahdi, Abdulla A., and Baraa T. Sharef. "Deep Learning Based An Optimized Predictive Academic Performance Approach." In *2023 International Conference on IT Innovation and Knowledge Discovery (ITIKD)*, pp. 1-6. IEEE, 2023.

LIST OF PUBLICATIONS

Papers Published/Communicated in International Journals:

[1] Rahul, Rahul Katarya, "Deep auto encoder based on a transient search capsule network for student performance prediction", Multimedia Tools and Applications, Springer, IF- 3.6, SCIE (Published), DOI: https://link.springer.com/article/10.1007/s11042-022-14083-5.

[2] Rahul, Rahul Katarya, "A Systematic Review on Predicting the Performance of Students in Higher Education in Offline Mode Using Machine Learning Techniques", Wireless Personal Communications, Springer, IF- 2.2, SCIE (Published), DOI: https://link.springer.com/article/10.1007/s11277-023-10838-x.

[3] Rahul, Rahul Katarya, "Optimized Deep Learning Based Students Performance Analysis Based on the Influence of Social Media", Earth Science Informatics, Springer, IF- 2.7, SCIE (Communicated).

[4] Rahul, Rahul Katarya, "Discriminability Enhanced Transformer Architecture for Students Performance Prediction Using Enhanced Features", Applied Intelligence, Springer, IF- 3.4, SCIE (Communicated).

Papers Published in International Conferences:

[5] Rahul, Rahul Katarya published the research paper entitled "Impact of Supervised Classification Techniques for the Prediction of Student's Performance", 2020 4th IEEE International Conference on I-SMAC (IOT in Social, Mobile, Analytics and Cloud), Palladam, Tamil Nadu, India, 07-09 October, 2020 (Scopus Indexed) (Presented and Published on IEEE Xplore Digital Library), DOI: https://ieeexplore.ieee.org/abstract/document/9243360.

[6] Rahul, Rahul Katarya published the research paper entitled " A Review: Predicting the Performance of Students Using Machine Learning Classification Techniques", 3rd IEEE International Conference on I-SMAC (IOT in Social, Mobile, Analytics and Cloud), Palladam, Tamil Nadu, India, 11-13 December, 2019 (Scopus Indexed) (Presented and Published on IEEE Xplore Digital Library), DOI: https://ieeexplore.ieee.org/abstract/document/9032493.

[7] Rahul, Rahul Katarya accepted the research paper entitled " Shapley Explainable Deep Learning based Knowledge Distillation Framework for Student's Performance Prediction ", 2nd IEEE International Conference on Data Science and Information System, Hassan, India, 17-18 May, 2024 (Scopus Indexed) (Presented and Published on IEEE Xplore Digital Library), DOI: https://ieeexplore.ieee.org/abstract/document/10594386