MASTER OF SCIENCE (BIOTECHNOLOGY) [SONIYA] 2023

A STUDY ON PARADIGM SHIFT IN DIAGNOSING DISEASES USING ARTIFICIAL INTELLIGENCE: CASE STUDY ON POLYCYSTIC OVARIAN SYNDROME

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

MASTER OF SCIENCE

IN

BIOTECHNOLOGY

Submitted by:

SONIYA

2K21/MSCBIO/50

Under the supervision of:

Prof. Yasha Hasija

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

CANDIDATE'S DECLARATION

I Soniya, Roll Number: 2K21/MSCBIO/50, student of M.Sc. Biotechnology, hereby declare that the work which is presented in the Major Project entitled **— A Study of paradigm shift in diagnosing diseases using Artificial Intelligence: Case study on Polycystic Ovarian Syndrome** in partial fulfillment of the requirement for the award of the degree of Master of Science in Biotechnology and submitted to the Department of Biotechnology, Delhi Technological University, Delhi, is an authentic record of my own carried out work during the period from January-May 2023, under the supervision of Prof. Yasha Hasija. The matter presented in this report has not been submitted by me for the award for any other degree of this or any other Institute/University.

The following attached work has been accepted in IEEE conference with the following details:

Title of the Paper: In-silico approach to DNA codon optimization of Zaire Ebola virus genes for vaccine development

Author Names: Soniya, Tirkey, Ruchi and Hasija, Yasha

Name of Conference*:* The International Conference for Convergence in Technology (I2CT)

Conference Date and Venue: 7^{th}-9^{th} April 2023/ Venue- Hotel Vivanta Pune, Hinjawadi, Maharshtra, India

Registration: Done

Status of Paper (Accepted/Communicated/Published): Published

Date of Paper Communication: 28 February, 2023

Date of Paper Acceptance: 12 March, 2023

Date of Paper Publication: 23 May, 2023

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi - 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "**Application Of Deep Learning In Diagnosis Of Polycystic Ovarian Syndrome (PCOS) Using Ultrasound Images Data**" which is submitted by **Soniya** (**2K21/MSCBIO/50**), Department of Biotechnology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Science is recorded for the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any degree or any diploma to this university or elsewhere.

Place: Delhi
Date:

Prof. Yasha Hasija
(Supervisor)
Professor
Department of Biotechnology
Delhi Technological University

Prof. Pravir Kumar
Head of Department

Department of Biotechnology
Delhi Technological University

ACKNOWLEDGEMENT

This is my heartiest gratitude to my supervisor Dr. Yasha Hasija, who has never failed to provide me all the best opportunities and resources for the successful completion of this dissertation. This research has been done in the field of my utmost interest, to which she has been a guiding light throughout the period of six months. A source of inspiration, Dr. Yasha Hasija, has showered me with wisdom and dynamism.

Sincere gratitude to Dr. Pravir Kumar, Head of Department of Biotechnology, DTU, who provided all the necessary resources required for the completion of this project work.

Big thanks to my research mate, Ruchi Tirkey, along with whom I share best of my project memories. It was her, who motivated me in tough times and we were able to submit conference paper in IEEE.

Special thanks to Mrs Jayshree Meena, Miss Neha Kumari for bein the wonderful mentors and helping me with all the knowledge they have.

The lab staff of department of biotechnology has always had my back, and I am blessed to have their constant support and help whenever I needed it.

All my classmates, including Bharmjeet Dahiya, Manju Dahiya, Tanya Srivastava, and lab mates of Computational Biology lab have been nothing less than a blessing who have helped me whenever I stuck.

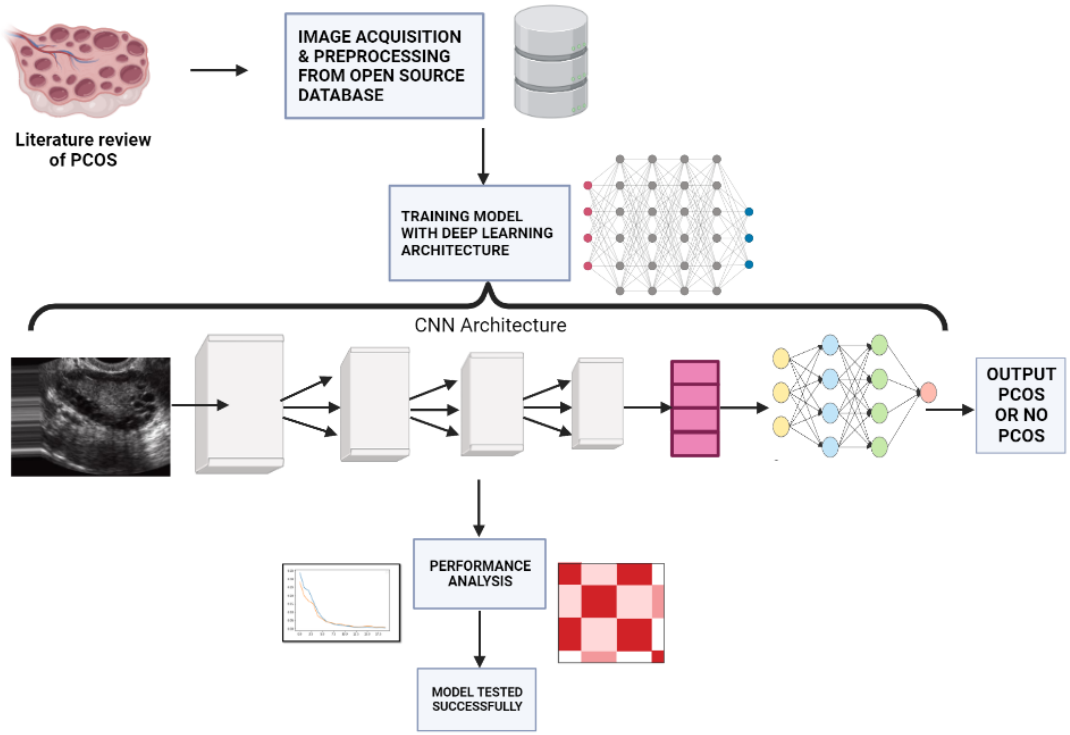
Last but not least, I acknowledge the support, care and love given by my parents and family. Without them, I won't be here.

SONIYA
2K21/MSCBIO/50

ABSTRACT

Polycystic Ovarian Syndrome is a complicated multifactorial disease identified as irregular menses, stubborn weight gain, insulin resistance, Hair loss, and in extreme cases, infertility and ovarian cancer. There are multiple diagnostic criteria but none of them covers all the aspects of PCOS and hence it becomes all the more important to perform absolutely correct diagnosis. In this case, Artificial intelligence has significantly played an important role. The present study discusses CNN-Convolutional Neural Network model using ultrasound images of infected and normal ovaries. The dataset was obtained from Open source database- Kaggle. With an accuracy of 74.5%, the model worked satisfactorily. To analyze the model, a normalized confusion matrix was also prepared, to rule out any confusion faced by model. Since the CNN Architecture was pretty straightforward and simple, accuracy was limited to 75%, however, with addition of more model layers, the accuracy can be increased. The amount of dataset can also be increased to improve efficiency of the model. With time, more techniques have been developed that can enhance the quality of healthcare in world. Smart AI like XAI or explainable AI is self-interpretable and helps working with complex algorithms. Precision medicine is also developing and innovative self-diagnosing applications can be built for diseases like PCOS which progress on symptoms that are easily identifiable and non-invasive.

GRAPHICAL ABSTRACT



Contents

LIST OF FIGURES: 9
LIST OF ABBREVIATIONS 10
CHAPTER 1 12
INTRODUCTION: 12
CHAPTER 2 15
LITERATURE REVIEW 15
2.1 UNDERSTANDING PCOS 15
2.1.1. AETIOLOGY OF PCOS 16
2.1.2. PCOS & INSULIN RESISTANCE 16
2.1.3. PCOS & HYPERANDROGENISM 17
2.1.4. PCOS & DISTURBED FOLLICULAR DEVELOPMENT 18
2.1.5. PCOS & ENVIRONMENTAL FACTORS- DIETARY IMABALANCE 18
2.2. DIAGNOSIS OF PCOS 19
2.3 TREATMENT STRATEGIES OF PCOS 22
2.3.1. TREATING HYPERANDROGENISM: 22
2.3.2. TACKLING HYPERINSULINEMIA: 22
2.3.3. MANAGING OTHER SYMPTOMS OF PCOS: 22
CHAPTER 3 24
3.2. MACHINE LEARNING 25
3.2.1. SUPERVISED LEARNING: 25
3.2.2. UNSUPERVISED MACHINE LEARNING: 26
3.2.3. SEMI SUPERVISED LEARNING: 26
3.2.4. REINFORCEMENT LEARNING: 26
3.3. NEURAL NETWORK AND DEEP LEARNING: 26
3.3.1. FEEDFORWARD NEURAL NETWORK 27
3.3.2. CONVOLUTIONAL NEURAL NETWORK 28
CHAPTER 4 30
METHODOLOGY 30
4.1. DATA RETRIEVAL 30
4.2. IMAGE PROCESSING AND AUGMENTATION 30
4.3. APPLYING CNN MODEL TO CLASSIFY ULTRASOUND IMAGES OF FEMALE OVARIES WITH AND WITHOUT PCOS 31

CHAPTER-5 34
RESULTS: 34
5.2. RESULT ANALYSIS 37
5.2.1. INFECTED Vs. NORMAL OVARY: 37
5.2.2. TESTING MODEL TO PREDICT WITH A TEST ULTRASOUND IMAGE OF AN INFECTED OVARY 37
5.2.3. CONFUSION MATRIX 38
CHAPTER-6 41
CONCLUSION 41
CHAPTER – 7 43
FUTURE PROSPECTS OF ARTIFICIAL INTELLIGENCE IN PCOS 43
REFERENCES: 45
APPENDIX 51
A.2 Code for using categorical cross entropy to calculate losses in model. 51
A.3 CNN Model using sequential() function showing trainable parameters 52
A.4 Code for preparing Confusion Matrix 52
PROOF OF CONFERENCE 53

LIST OF FIGURES:

Figure number	Figure caption
Figure 2.1	Number of randomized controlled trials on Dietary effects on PCOS available in last 5 years on Pubmed
Figure 2.2	Interplay of genetic, environmental, and hormonal factors in pathophysiology of Polycystic Ovarian Syndrome
Figure 2.3	Schematic diagram representing diagnostic criteria of Polycystic ovarian syndrome and the possible phenotypes.
Figure 3.1	Increase in number of peer reviewed articles on applications of AI in healthcare sector.
Figure 3.2	Relation between AI and its subgroups; ML & DL
Figure 3.3	A) Single perceptron. B) Network of perceptron in multiple layers forming an Artificial Neural network with information moving in forward direction (Feedforward neural network).
Figure 4.1	Augmentation of image datasets to increase the efficiency of model
Figure 4.2	Convolutional Neural Network employed in Detection of PCOS using Ultrasound images of ovaries
Figure 4.3	CNN Models deployed in the study: Convolutional layer, Pooling, Flatten and Dense.
Figure 5.1	Overall Model accuracy is found to be 74.5%
Figure 5.2	With each Epoch, training loss and validation loss is decreasing which explains how well the model is fitting our data, and vice versa, respectively.
Figure 5.3	With each Epoch, training accuracy and validation accuracy is increasing which explains how well the the model is predicting the unseen validation data.
Figure 5.4	Difference in normal and infected ovaries with PCOS as visible in ultrasound images of ovaries.
Figure 5.5	Input Code to extract test image and check for model's prediction
Figure 5.6	Input image: Polycystic ovary
Figure 5.7	Prediction result showing diagnosis as: INFECTED
Figure 5.8	A sample confusion matrix
Figure 5.9	Number of true positive (TP) and true negative (TN) are shown by light blue colored boxes whereas False Positive and False negatives are shown by Dark blue colored boxes.
Figure 7.1	The future of Artificial Intelligence will require a strong interrelation of each of the elements: Big Data, Hardware and Algorithm to produce Smart solutions
Figure A.1	Code for using categorical cross entropy to calculate losses in model

Figure A.2	CNN Model using sequential() function showing trainable parameters
Figure A.3	Code for preparing Confusion Matrix

LIST OF ABBREVIATIONS

S.No	ABBREVIATION	FULL FORM
1.	**AI**	Artificial Intelligence
2.	**ML**	Machine Learning
3.	**DL**	Deep Learning
4.	**PCOS**	Polycystic Ovarian Syndrome
5.	**HA**	Hyperandrogenism
6.	**IR**	Insulin Resistance
5.	**CYP450**	Cytochrome P 450
6.	**T2D**	Type-II Diabetes
7.	**IGT**	Impaired Glucose Tolerance
8.	**MTNRB1**	Melatonin Receptor B1
9.	**AMH**	Anti-mullerin hormone
10.	**DHEA**	Dehydroepiandrosterone
11.	**SHBG**	Sex Hormone Binding Globulin
12.	**ACTH**	Adrenocorticotropic hormone
13.	**PA**	Pre-natal Androgen
14.	**GWAS**	Genome Wide association linkage
15.	**ITGAX**	Integrin alpha x
16.	**ET-1**	Endothelin-1
17.	**TNF**	Tumor necrosis factor
18.	**NO**	Nitric oxide
19.	**LH**	Leutinizing Hormone
20.	**FSH**	Follicular stimulating Hormone
21.	**IGF**	Insulin like growth factor
22.	**HI**	Hyperinsulinemia
23.	**OD**	Ovulatory dysfunction
24.	**AES**	Androgen excess society
25.	**NIH**	National Institute of Health
26.	**PCOM**	Polycystic ovarian Morphology
27.	**FFA**	Free fatty acids
28.	**ANN**	Artificial neural network
29.	**SML**	Supervised machine learning
30.	**ReLU**	Rectified Linear Unit
31.	**XAI**	Explainable AI
32.	**GnRH**	Gonadotropin hormone releasing hormone
33.	**GABA**	Gamma- aminobutyric acid

34.	**CBSF**	Cerebospinal Fluid
35.	**CREBZF**	Camp- response element binding protein/ATF BZIP
36.	**HSD**	Hydroxysteroid dehydrogenase
37.	**AMPK**	AMP- activated protein kinase

CHAPTER 1

INTRODUCTION:

One in every 15 women globally suffers from polycystic ovary syndrome (PCOS), an endocrine condition. It is primarily characterized by elevated levels of androgens, acne, and hirsutism, and can eventually make women more susceptible to long-term insulin resistance, miscarriages, or even infertility. It is a chronic disease and has been well documented in all stages of women life. Symptomatic management has led us to a state where PCOS can be controlled but till now, a lot of females face issues in conceiving due to PCOS. Despite being recognized for several decades, PCOS remains a complex and challenging condition to understand fully [1]. The etiology of subject disease is definitely complex and involves delicate relation between hormonal, genetic, metabolic, and environmental factors like dietary choices. Genetic predisposition appears to play a role, as PCOS often clusters within families. Various candidate genes have been identified, suggesting a complex polygenic basis for the syndrome [2]. Hormone imbalances are the most common causative or diagnostic markers of PCOD, for example excess androgen secretion or Hyperandrogenism (HA) or disregulated gonadotropin secretion, Insulin resistance or hyperinsulinemia are also significant in PCOS [3]. Multiple environmental factors are also responsible for making PCOS a lifestyle disorder as well. Unhealthy diet choices like high fat, low protein intake along with lethargic schedules with less or no day-to-day activity is disturbing the lipid profile of body leading to obesity or adipose accumulation resulting in aggravated PCOS symptoms [4].

To this date, there is no cure for PCOS as unfortunately, it is a chronic disorder. Every patient has a different symptom, some are dealing with T2D, while others are facing problems in conceiving. This is one of the major reasons that a single cure can't be decoded for PCOS. However recent advancements in healthcare give promising hopes that can change the future of disease management in world.

Artificial intelligence is a computer science based contemporary technique that creates programs and algorithms to make machines smart and effective at carrying out tasks that often call for expert human intellect. It was introduced in 1950s by Alan Turing's

attempt to distinguish whether a machine can be intelligent as a man via 'The Turing Test'. But it took 20 years for AI to enter into the Healthcare sector with the onset of databases like PubMed and MEDLARS that digitized the data. The digitization framed a strong foundation for Artificial Intelligence in Medicine (AIM). From MYCIN (that could recommend antibiotic treatment options according to a patient's body weight) to using new technologies to identify ALS, there is no looking back for AIM. Recently due to advanced research in Artificial Intelligence, the applications in healthcare industry have increased exponentially [5]. It has applications in prediction, diagnosis, treatment and even management of diseases due to the huge amount of healthcare data available to us.

As a subgroup of artificial intelligence, Machine learning works as a human brain in a machine. A code is trained on the basis of a large amount of data to distinguish or make a decision on specific question which is input in different forms, be it an image, a text or numeral based data and even voice [6]. In the present study, we will understand how machine learning has mediated in shifting healthcare from invasive diagnostic methods that are painful to the patients to a noninvasive way of diagnosing disorders like PCOS using ultrasound images. The present study utilizes Convolutional Neural Network that specializes in processing data which has a grid like architecture, for example an image. To define an image, it is a binary representation of visible data. Each grid contains pixels that represent how bright and which color the pixel should be [7]. A.K.M Hosain et al built PCONet which is a CNN model to detect or diagnose PCOS with an accuracy of 98% [8]. Various other similar studies have been done to increase the accuracy of detection without any error. Chauhan et al. used Naïve Bayes, Decision tree classifier, k-nearest algorithm aka KNN, SVM & LR to detect Polycystic ovaries where Decision tree defeated all the other. On the basis of a picture dataset, Kumari categorised PCOS using diverse ML algorithms. According to her findings, "VGG-19" had the best accuracy out of the four, scoring at 70%. By employing the Generative Adversarial Networks (GANs) and enhancing the data, she managed to get over the limitations imposed by the dataset. In total, there were 94 images; 50 of them showed PCOS, while the other 44 were ultrasonic pictures that didn't [9].

Objectives of the study;

1. Understanding the employment of Artificial intelligence in diagnosing diseases like PCOS.
2. Developing a CNN model using ultrasound image dataset available in the open source database – Kaggle
3. Analyze the precision of the proposed CNN model with the help of Confusion matrix.
4. Propose future prospects of artificial intelligence in management of PCOS.

CHAPTER 2

LITERATURE REVIEW

2.1 UNDERSTANDING PCOS

Polycystic Ovarian Syndrome (PCOS) is known to be the most common disorder in women of 14-49 years of age. A patient may present asymptomatically or with several gynecologic, dermatologic, or metabolic symptoms. The Rotterdam criteria, asks for only two of these 3 symptoms— HA, OD, & ovaries with multiple cysts— along with discarding other disorders that show HA or ovulatory dysfunction, are cordially suggested by the Endocrine Society's guidelines. Menstrual abnormalities, abnormal hair growth (hirsutism), overweight and obesity, Type 2 diabetes, and heart disease are all prevalent among those with PCOS. Not just physiological, but PCOS also causes mental dysfunctions in multiple patients, for example women face depression, anxiety and hormone imbalance while dealing with irregular periods, cases of infertility, and other symptoms as stated above [10].

Due to multiple common symptoms, this heterogenous disorder is also a challenge to diagnose. Detecting Hyperandrogenism is relatively easy by identifying an increased level of androgen in blood tests but can give a false positive result as well. Presence of cysts in ovaries is diagnosed via ultrasound imaging which can also help to check the lining of endometrium. However, since the counting of cysts can be an error prone process and might present us with false diagnosis Irregular periods will lead to thickening of endometrium, this can also lead to Uterine Cancer. Therefore, it is unnecessary to mention that PCOS needs more attention than it is given. Up until now, there are only symptom based treatments available for example, in case of hirturism, first line of drugs include Spironolactone, Metformin and Eflornithine. However for other effects like infertility, clomiphene is given. Because of high cases of Insulin resistance in PCOS patients, Metformin is most commonly used drug [11].

2.1.1. AETIOLOGY OF PCOS

To understand PCOS, let us understand the causative mechanisms of Polycystic Ovarian Syndrome. One in every 15 women globally suffers from polycystic ovary syndrome (PCOS), an endocrine condition. It is primarily characterized by elevated levels of androgens, acne, and hirsutism, and can eventually make women more susceptible to long-term insulin resistance, miscarriages, or even infertility. It is a chronic disease and has been well studied and reported at every stage of women life after menarche. Most women with PCOS are obese, particularly at abdomen which puts them at a higher risk of developing secondary diseases like Impaired Glucose Tolerance (IGT), Type 2 Diabetes Mellitus, stress, anxiety, Pancreatic cell dysfunction and other metabolic problems [1]. PCOS can be caused because of genetic as well as Environmental factors as demonstrated in Fig 2.2. After years of study, studies could finally jot down some significant causes of PCOS, some of them are mentioned: Reactive oxygen species (ROS), increased exposure of androgen to embryo, immunological and endocrine complications. It is also affected simultaneously by many genes or oligomeric sequences. "genome-wide association studies (GWAS)," genes associated with specific location, and foetal programs revealed the genetic basis of HA and IR and the probable involvement of acquired factors from the environment in PCOS. [12]

2.1.2. PCOS & INSULIN RESISTANCE

International Diabetes Federation released a statement described as follows; PCOS is a risk factor for T2D which cannot be modified or prevented, especially in women. IR (Insulin Resistance) acts as a shared link between PCOS and T2D. Even after IR being a complex situation, blood lineage with history of IR and obesity have been found to be common in PCOS affected women. This means families or parents can pass PCOS to their off springs [2][3]. Wonem with PCOS show increased levels of phosphorylated receptor substrata that is the cause of feeble signal from insulin receptor. Mutation in Melatonin Receptor B1 (MTNRB1) is found in PCOS which retards the insulin production, accumulating glucose. Women with Insulin resistance are more likely to develop PCOS because insulin stimulates theca cells to produce androgen, reduces free testosterone, ultimately results in down regulation of SHBG in liver. The aberrant steroidal function in the ovary, increased exercise by P450 steroidogenal enzyme associated with production of male hormones, problems with cortisol metabolism are some other possible causes of elevated androgen levels. Circulating ACTH levels in PCOS-afflicted women are found to be comparable to those in controls. For babies

exposed to PA, the unwanted androgen increase for effects of glucose digestion and precursors for upcoming body illnesses is substantially more. Adipocyte growth may also be promoted by testosterone [13]. Infants with PA experience a drop in average islet size, proportional islet growth, and stable islet fractional area. Additionally, such newborns also demonstrated that the islet cell ratio was rising and the growth marker Ki67 was elevated [14].

IR leads to Hyperandrogenism but is unrelated to weight gain in PCOS patients. In women with similar weights, Insulin mediated glucose metabolism was reduced to a certain 40% in women with PCOS than in healthy patients. Obesity might not cause this reduction however, it can aid indirectly [15]. PCOS patients represent higher levels of CD11c (ITGAX) and TNF-alpha gene expression in cutaneous adipose tissue than the general population. TNF being the inflammatory agent can boost the development of IR in PCOS women. Endothelial artery cells develop Insulin resistance when NO level falls and ET-1 levels rise. As a result, vasodilation via insulin is restricted due to decrease in NO production and high vasoconstrictors production. The American Heart Association therefore concludes that IR accelerates the risk of enzyme involved and cardiovascular diseases in PCOS-positive individuals along with T2D. In PCOS women, IR leads to an increase in insulin release from pancreatic cells. IR also elevates hepatic production, mobilizes adipose tissues to increase quantity of plasma-FFA. It inactivates enzymes including PDH and ultimately leads to glucose production and insulin resistance [16].

2.1.3. PCOS & HYPERANDROGENISM

HA or Hyperandrogenism or increased/unwanted expression of androgen in female body, is not only the main diagnostic marker of PCOS but also a major causative condition of PCOS. If mother's uterus reports a high level of androgen and preterm delivery, fetus is most likely to develop PCOS [17]. Daughters of PCOS afflicted mother secrete higher anti-mullerian hormone (AMH) that slows down the initial follicular growth and contributes in accumulation of follicles in ovaries referred to as a polycystic ovary. Adrenarche begins at an early stage which is the greatest trouble in PCOS [18].

2.1.4. PCOS & DISTURBED FOLLICULAR DEVELOPMENT

PCOS develops in the adolescent years of life where symptoms like acne, irregular menses, physical disturbances due to increased levels of luteinizing hormone, LH/FSH ratio, testosterone, DHEA, simultaneous reduction of SHBG along with IGF binding protein are reported due to Hyperinsulinemia and IR. HI induces release of LH from pituitary, thus LH/FSH ratio increases, increases androgen synthesis from ovaries, downregulates SHBG which in turn leaves free testosterone as mentioned above. High levels of ROS, chemokine and cytokine production in response to impaired adipose tissues causes inflammation which is reported to be one of the causes of HA [15]. All in all, HA along with augmented levels of LH muddles the ovulatory maturation process and lead to anovulatory (no egg) cycles- infertility.

2.1.5. PCOS & ENVIRONMENTAL FACTORS- DIETARY IMABALANCE

More recently, PCOS has been identified as a lifestyle disorder where most of the symptoms can be treated if patient is given a balanced diet. Obesity is prevalent in 35-70% of PCOS patients and weight gain simply worsens the situation while a healthy weight loss can improve it. There is not enough literature backed data on effects of diet regulation on PCOS. Fig 1 depicts the number of randomized controlled trials in last 5 years when searched on PubMed with keywords ‘PCOS & Diet’.

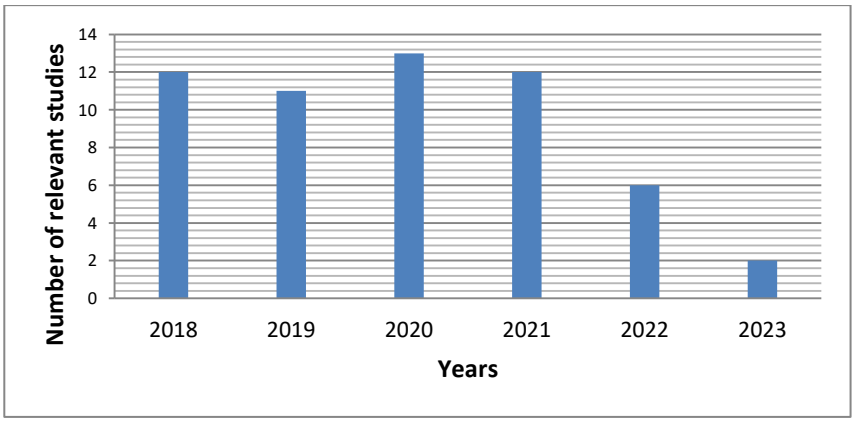


Fig 2.1: Number of randomized controlled trials on Dietary effects on PCOS available in last 5 years on Pubmed

Women with higher fats intake develop severe complications in PCOS. Carbohydrates and protein levels are also studied in relation to PCOS. Low carbohydrates in diet reduce IR and cholesterol levels, along with overall reduction in weight up to 5% [4].

Low glycemic load, high protein and low caloric diet have been reported to increase insulin sensitivity to a significant level. An overall weight loss of 5-10% can help with relieving the distress, according to recent guidelines, however lower energy diets have little sustainability, and weight management is a task to fulfill for long term. Even if the patients are highly motivated to lose weight, they are heavier than normal individuals, this is a changed response to diet in affected female's body, may be because of HA and IR, having a role in changing appetite, metabolism and energy homeostasis.

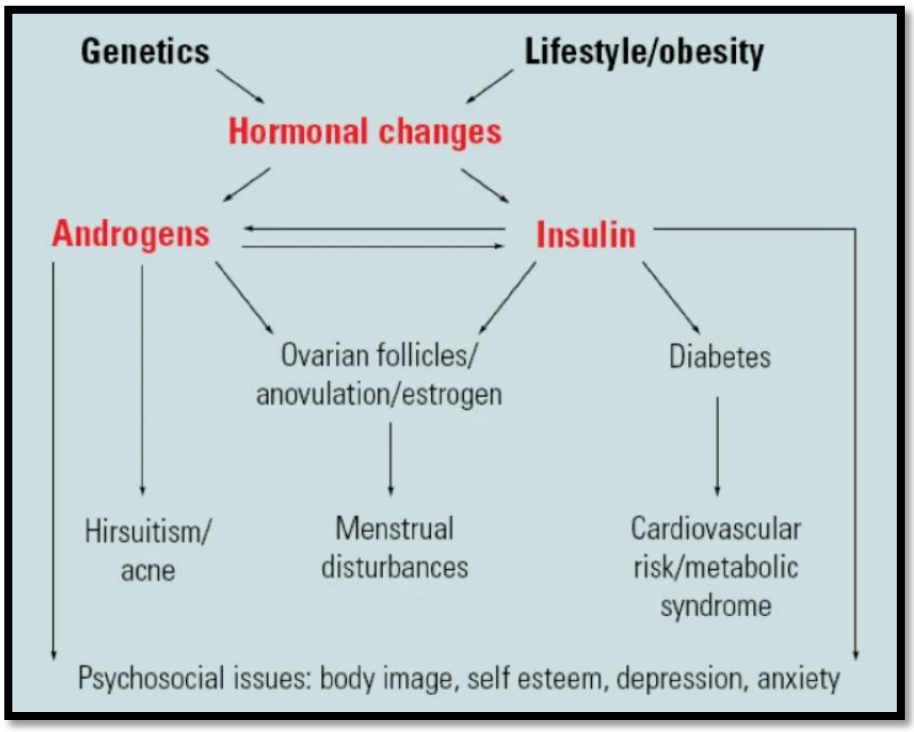


Fig 2.2: Interplay of genetic, environmental, and hormonal factors in pathophysiology of Polycystic Ovarian Syndrome

2.2. DIAGNOSIS OF PCOS

Even after being prevalent in human population since decades, PCOS is not given the attention it screams for. Until 2018, there were no consistent guidelines to characterize a polycystic ovary with other ovarian conditions correctly[19]. In 1930s, Steien & Levanthall were the first ones to report a syndrome with menstrual irregularity, hirsutism and infertility. It was then called Steien- Levanthall syndrome and later was renamed as PCOS. Until 1970s, PCOS remained a reproductive disorder, but with the discovery of conditions like insulin resistance, HA and hirsutism caused by mutations in Insulin receptor, they were associated with each other.

PCOS is determined by a number of markers which is both a good and a bad thing. Since there are a multiple markers, multiple detection routes are available but they may

also be linked to some other condition and give false positive as PCOS. Increased levels of LH wrt FSH. This LH: FSH ratio can abscond in random blood tests due to fluctuating LH release [20]. High level of free testosterone is significantly reported in PCOS patients. In cases of PCOS that leads to anovulatary cycles, reduced progesterone along with estradiol in mid-follicular level is established through studies [21]. Ultrasound images characterize PCOS by an augmented number of antral follicles, thickening of ovarian borders and a significantly visible hyperplasia in theca cells. This induces production of more androgen in response to LH from the theca cells. PCOM (polycystic ovarian morphology; One ovary containing more than 12 follicles) is most common in women with regular menstrual cycle and decreases with increase in age. Until 8 years of menarche, it is not recommended as a diagnostic feature for PCOS. Some women with PCOM have higher levels of testosterone, but others might have normal levels. A study showed PCOM affected women never develop PCOS during follow up [22].

As it is clear that the causes of PCOS is not very clear, the "vicious cycle of PCOS" has not been decoded, so the diagnosis is based on phenotypic characters of disease [23].

National Institute of health (NIH) criterion was adopted in 1990 at an international conference meeting where every participant decided/participated for the potential diagnosis factor. HA and chronic anovulation received maximum votes and came to be known as NIH Criteria [24].

2.2.1. Rotterdam criteria was adopted in 2003 in Netherlands. It was a very random conference discussing PCOS diagnosis, and the finalized criterion was not evidence based. It was just an add on to NIH criteria. Rotterdam required any two of the three symptoms to be present for PCOS diagnosis

1) Hyperandrogenism,
2) Chronic anovulation and
3) PCOM

It broadened the affected groups of women as

1) PCOM with HA and no Ovulatory dysfunction
2) PCOM with Ovulatory dysfunction and no HA [25]

2.2.2. Androgen Excess society gave AES criteria in 2006 with the help of a dedicated task force who gave evidence based suggestions.

Clinical/Biochemical HA with

1) Ovarian dysfunction with or without PCOM

2) PCOM minus OD

This has not been widely adopted and it also stamped out the Rotterdam criteria [26]

2.2.3. NIH worked on evidence based diagnosis identification on PCOS in 2012. The accentuation on PCOM caused perplexity because it wasn't obligatory or ample for the diagnosis of PCOS, according to the conference group's final report, and the word "PCOS" was a hindrance to advancement. The panel proceeded by utilizing the Rotterdam criteria in investigations with exact phenotypic definitions. Additionally, they suggested a thorough study plan that includes evaluation of the physiology and chronic health effects of the PCOS symptoms.

2.2.4. The 2013 Clinical Practice Guideline of the Endocrine Society and the **International Evidence-Based Guideline of 2018** are the 2 main evidence-based recommendations for the detection & superintendence of PCOS, respectively [27]. Because of the dearth of randomised clinical trials (RCTs) in the PCOS, standard of proof on which these criteria are founded or constructed is mostly low. The evidence supporting twenty fours of the thirty four suggestions in the Endocrine Society Clinical Practise Guideline was deemed to be of poor or very low quality. Out of 175 recommendations, only 31 of them were adopted that were evidence backed. Both the guidelines, supported by experienced personalities, prefer Rotterdam criteria for high accuracy diagnosis of PCOS.

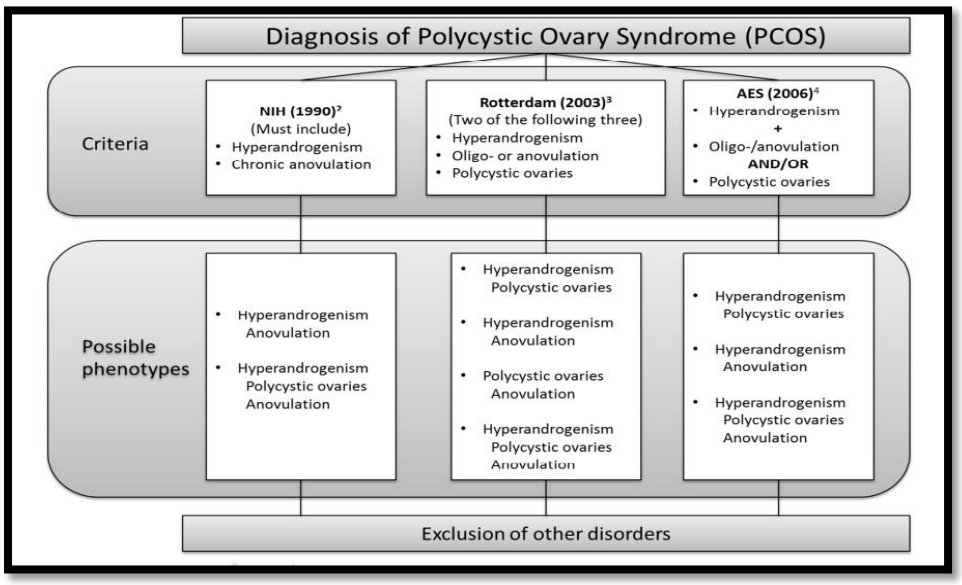


Fig 2.3: Schematic diagram representing diagnostic criteria of Polycystic ovarian syndrome and the possible phenotypes

2.3 TREATMENT STRATEGIES OF PCOS

2.3.1. TREATING HYPERANDROGENISM:

The main objective while treating a PCOS patient is to reduce the roaming free testosterone which gets accumulated due to inactive SHBG. To achieve this, there are some strategies like- Muting testosterone releasing enzymes, altering the perceptivity of theca cells that produce excessive androgen in response to LH, attenuating gonadotropin receptor hormone (GnRH) and AMH levels. GABA has been reported in high concentrations in CBSF of PCOS afflicted women which has been correlated to high levels of GnRH. In studies, women reported requirement of high levels of progesterone to control Leutinising Hormone levels [28]. Kisspeptin with an unknown origin has been reported to have a positive relation with LH in PCOS women. CREBZF gene has been reported to regulate testosterone and enzymes required in production of androgen. Gene silencing can be implicated to silence CREBZF gene. A natural compound called neringenin (flavanone) has been reported to inhibit activity of 3-beta HSD & 17 beta-HSD – the steroid enzymes [29]. Androstenedione is converted to testosterone via AKR – 1C3 enzyme and it is present in excess levels in PCOS afflicted women which can also be targeted as a therapy [30].

2.3.2. TACKLING HYPERINSULINEMIA:

Metformin is the holy grail for PCOS patients apart from clomiphene. Metformin acts as the activator of AMPK Pathways via inhibition of liver's respiratory chain [31]. The AMPK pathway in turn alters the fat metabolism pathway and leads to sensitization of insulin. It has been used in PCOS since more than 20 years and has not disappointed in managing it. It also inhibits fructose- 1,6 bisphosphate [32]. With research, it has been proven that metformin can induce ovulation in PCOS women with non-obese weight along with relieving acne, hair loss, etc.

2.3.3. MANAGING OTHER SYMPTOMS OF PCOS:

Aquaporins (7-9) have been working great as drug targets to grow normal follicles in PCOS affected women. Normal levels of Vit D also play extremely dominant part in preparing the endometrium. Females with normal levels of Vitamin D had more thick endometrium which is directly related to the chance of conceiving. Minodixil has been working fine in alopecia in PCOS. It acts as potassium channel opener and its

metabolite component minodixil sulfate is mainly responsible for treating alopecia. Hormonal imbalance is very frequent in PCOS, which can often lead to infertility in rare cases. This is the cause of anxiety disorders in many females with PCOS. With research, it can be concluded that healthy lifestyle and diet plays a major role is alleviating such symptoms. Drugs like clomiphene are available to induce ovulatory cycles [33].

CHAPTER 3

3.1. ARTIFICIAL INTELLIGENCE IN HEALTHCARE

The term was coined by John McCarthy in 1956, which denotes a machine which mimics human brain and its intelligence, including features like learning, decision making, and problem solving. Recently in last 10 years, researchers have started utilizing the full potential of Artificial intelligence. It is not just a single system, but a superset of multiple algorithms, one of them being Machine learning. The sudden success of AI and ML is attributed to the vast amount of digitized data that is now available. The only thing that can make or break a ML algorithm is the data [34].

Fig 3 depicts the sudden increase in number of articles published in last 10 years discussing applications of AI in healthcare sector.

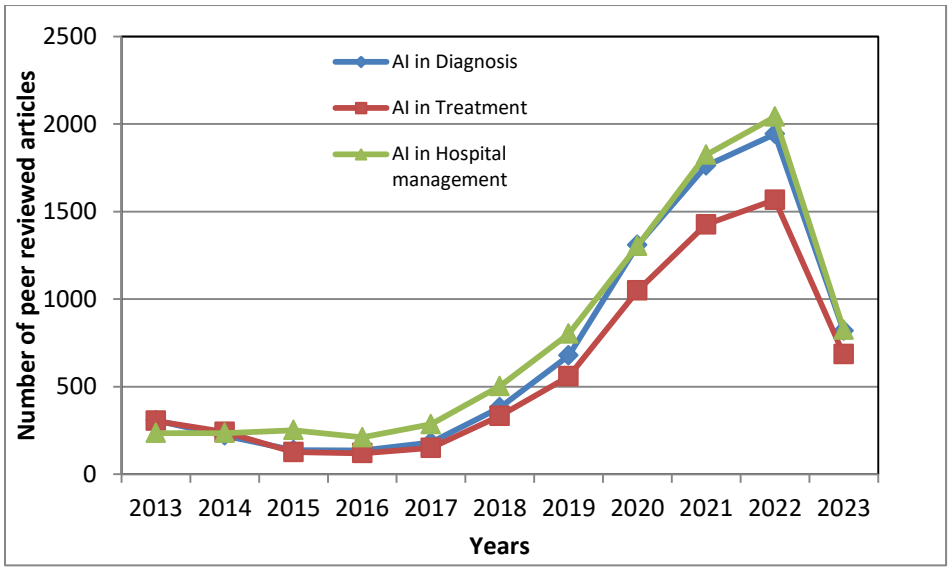


Fig 3.1: Increase in number of peer reviewed articles on applications of AI in healthcare sector.

3.2. MACHINE LEARNING

Unlike traditional programming, where an algorithm may be categorically programmed using well-known properties, machine learning (ML) takes into account subgroup of data to produce an system which is capable of using innovative or unconventional amalgamation of features rather than those that can be inferred from the ground up. The four most popular learning techniques in machine learning (ML)—as discussed following can each be used to tackle a distinct problem [35] .

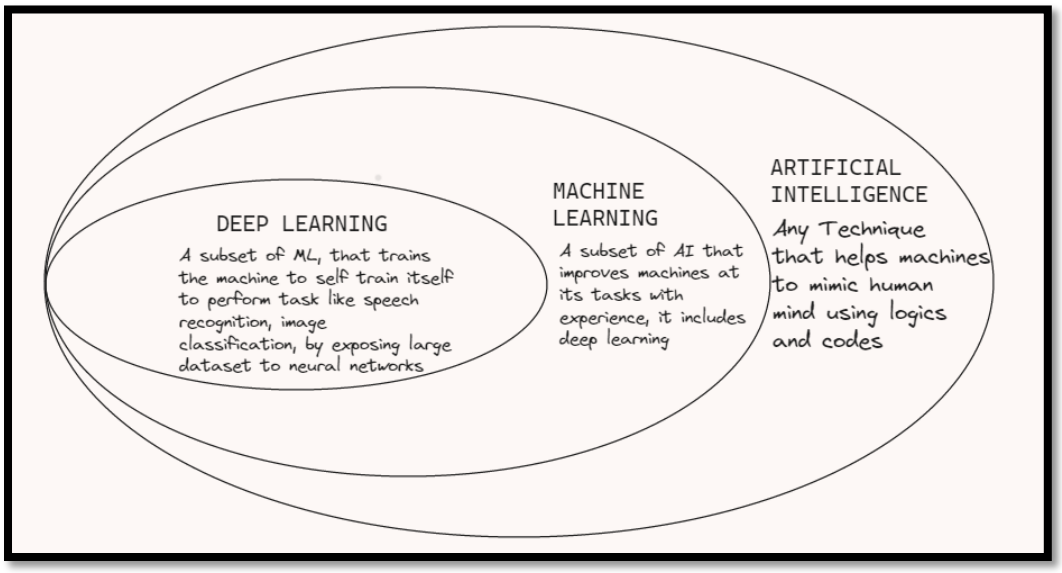


Fig 3.2: Schematic diagram showing relation between AI and its subgroups- ML & DL

3.2.1. SUPERVISED LEARNING:

It is a supervised training of machine where the training dataset is labeled explicitly. Following are the steps to perform supervised machine learning;

1. Data is possessed, split into training, testing and validating dataset
2. Inform the model about the link between features and target using the training dataset.
3. Gauge the model using test dataset to check how well it can predict the results [36]

Regression and classification are the two most commonly used SML algorithms, where regression predicts some absolute values like numeric data. However, a classification algorithm essentially classifies into categories to which the respective data belongs [36]

3.2.2. UNSUPERVISED MACHINE LEARNING:

Unsupervised learning seeks to identify specific designs in a datagroup and assign each instance to a certain category, in contrast to supervised learning. These algorithms are not superintended because the features in a dataset that exists or not are determined via the algorithm without any guidance using a labeled data or target [37]. Clustering and association are two of the unsupervised learning algorithms that are most frequently used [36] Clustering is simply dividing the data into specific clusters based on some patterns. As we go on preparing clusters, some unsupervised outputs present us with different features that we did not know of in the first place [36].

3.2.3. SEMI SUPERVISED LEARNING:

It is the middle situation between supervised and unsupervised learning. It comes in handy when dataset is not consistent, per se, some of it is labeled and rest is not. This type of algorithm is used in cases of image data when labeling each one of the image is time consuming and laborious. In these cases, an expert labels some images and then trains the model to label the rest of images. These labeled images then act as training set for another model that was actually needed for the problem [36]

3.2.4. REINFORCEMENT LEARNING:

This is the most advanced form of machine learning which does not solves problem on the basis of a labeled or unlabeled dataset. There are no desired outputs, like a human, it learns and attempts to answer using its experience along with a dedicate dataset. If it fails to provide appropriate answers, it learns again from its errors[38]. This is one of the closest algorithm, technology could get to human learning experience. Reinforcement learning has its applications in computer science but it is yet to prove its potential in Medicine [36].

3.3. NEURAL NETWORK AND DEEP LEARNING:

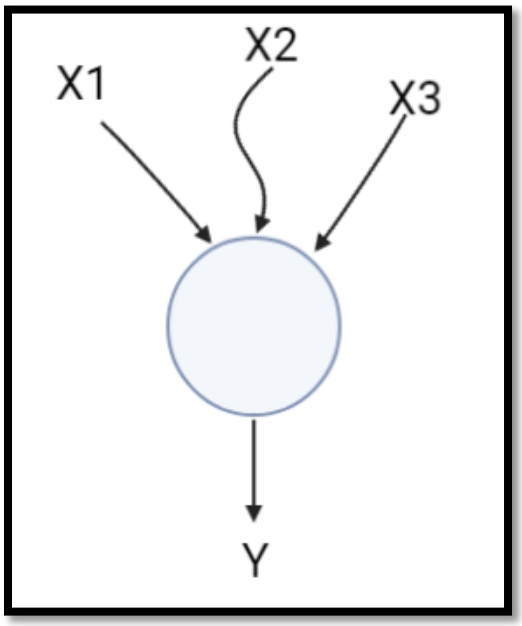
ANN is another subtype of Machine learning inspired from human neural networks. Nodes of ANN replicate Neurons in humans' body that interacts with each other. The strength of connections between nodes in an ANN is determined on the basis of their

capacity to produce an intended result, similar to human neuron system where neurons junctions are nourished more and more only when each nerve cell share equivalent response in a biological system [39].

3.3.1. FEEDFORWARD NEURAL NETWORK

As the name suggests, the information is transferred from one node of the earlier layer to the nodes present in the next consecutive layer, translated and then again transferred to the next node. The smallest unit of this model is a perceptron which is a ML model that takes in an input in form of important features and tries to identify a segregating line, hyper plane to classify them in respective tangents. Collectively, multiple perceptrons form multilayer perceptron model aka the Artificial neural network (ANN). When information is passed in forward direction in a perceptron model, it becomes a FeedForward neural network [39].

A)



B)

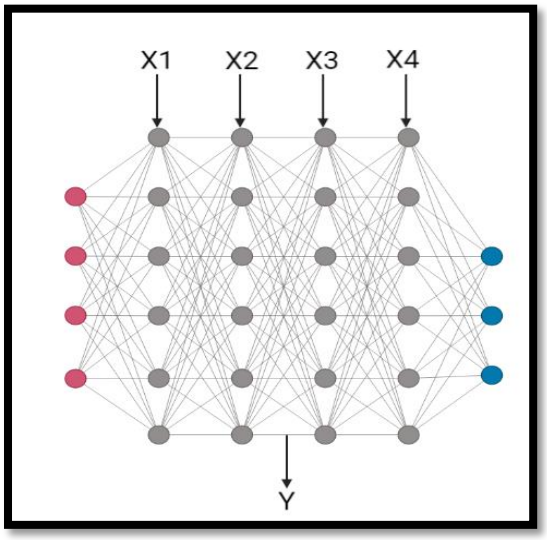


Fig 3.3: A) Single perceptron. B) Network of perceptron in multiple layers forming an Artificial Neural network with information moving in forward direction (Feedforward neural network).

3.3.2. CONVOLUTIONAL NEURAL NETWORK

In cases of image recognition, Feedforward ANN is not the ideal choice as pixels in an image are the input for each node and there is no as such relation between layers and nodes. This can distort the spatial arrangement of pixels. To solve this technical issue, CNN comes handy. Instead of taking single pixel as input, CNN puts batch of pixels into specific nodes in the next layer to preserve the spatial arrangement of image [40]. These batches of nodes absorb the special features that will be used as convolutional filters. Grey scale images are simply single matrix digital-image and 3 stacked matrix if colorful image (blue, red and green channels). Pixels or matrices have values ranging from 0 to 255 according to intensities of color [39] The convolutional filter is a little more smaller matrix of size two by two to Nine by nine. The filter matrix is multiplied with image matrix through which the convolutional filter passes and a new matrix is created that entails values which tell if the filter has caught a feature of interest [40].

The features are extracted and their location is marked on feature map. This is used as input for next layer via a deep CNN algorithm. The next layer produces new feature map using new filters and when this goes on for multiple layers, the final feature map is

extracted, compressed and input to ANN which ultimately performs image identification. All of this is known as Deep Learning [39].

Deep learning has been effectively used in exploiting image datasets of diverse patient and their affected regions. For example, a study by Wenqi et al. with an accuracy of 98%, CNN model trained to diagnose PCOS used scleral images [41]

DL has also shown applications in Diabetic Retinopathy using retinal fundus images [42]. The only challenge available to us right now is the amount of data for diseases like PCOS. Also, large data is not always the solution, if the training data is imperfect, mislabeled and unrelated to real world population or the test population, the model will fail miserably to perform well in a real life scenario. Another problem commonly referred to as “black box” is very frequent with DL algorithms. It is a condition when input is given and an output is received but on what basis the output came, is not very clear. In such cases, simpler linear algorithm is much more interpretable than DL [39].

CHAPTER 4

METHODOLOGY

4.1. DATA RETRIEVAL

Datasets have been taken from Kaggle, an online community for information scientists and AI experts. Customers can search and submit informational indexes on Kaggle, research and construct models in an electronic information science environment, and integrate with many information scientists that specialize in Artificial Intelligence who are motivated by challenges to complete information science assignments. It is a great learning platform along with an authentic source of information and datasets. A raw dataset with 1568 PCOS positive images of ultrasound, and 2288 not-infected ultrasound images of ovaries was obtained from an open source for public use from Kaggle.

A raw dataset with 1568 PCOS positive images of ultrasound, and 2288 not-infected ultrasound images of ovaries was obtained from an open source for public use from Kaggle.

4.2. IMAGE PROCESSING AND AUGMENTATION

As the images were not consistently of similar sizes and colors, they were not suitable to use in machine learning. Hence, after acquiring images, they are subjected to certain changes like resized all the images to 224 x 224, adjusted color to standard ‘RBG’. Jupyter Notebook has been utilized in this case as interface for exercising, primarily combining TensorFlow and Scikit-learning packages for Python.

Augmentation of images can be a very useful as well as quite powerful technique to reduce the prediction errors. The number of images is increased by altering the sharpness, enhancing image contrast, flipping, shifting etc. In this study, some manipulations are done including zoom, horizontal and vertical flip and rotating.

```
20]:
from tensorflow.keras.preprocessing.image import ImageDataGenerator
# create generator
datagen = ImageDataGenerator(rescale = 1./255,
                             shear_range = 0.2,
                             zoom_range = 0.2,
                             horizontal_flip = True,
                             vertical_flip=True,
                             rotation_range=30,
                             validation_split=0.3,
                             fill_mode='nearest'
                             )
# prepare an iterators for each dataset
train_it = datagen.flow_from_directory( directory,
                                        class_mode='categorical',
                                        classes=['infected', 'notinfected'],
                                        target_size=(224, 224),
                                        batch_size=100,
                                        subset='training',
                                        seed=24)
# prepare an iterators for each dataset
val_it = datagen.flow_from_directory( directory,
                                        class_mode='categorical',
                                        classes=['infected', 'notinfected'],
                                        target_size=(224, 224),
                                        batch_size=100,
                                        subset='validation',
                                        seed=24)
```

Fig 4.1: Augmentation of image datasets to increase the efficiency of model

4.3. APPLYING CNN MODEL TO CLASSIFY ULTRASOUND IMAGES OF FEMALE OVARIES WITH AND WITHOUT PCOS

Deep Neural Network architecture provides us with a very useful model called Convolutional Neural Network algorithm to classify images. Fig 6 depicts the architecture of CNN model employed for diagnosis of Polycystic ovarian syndrome.

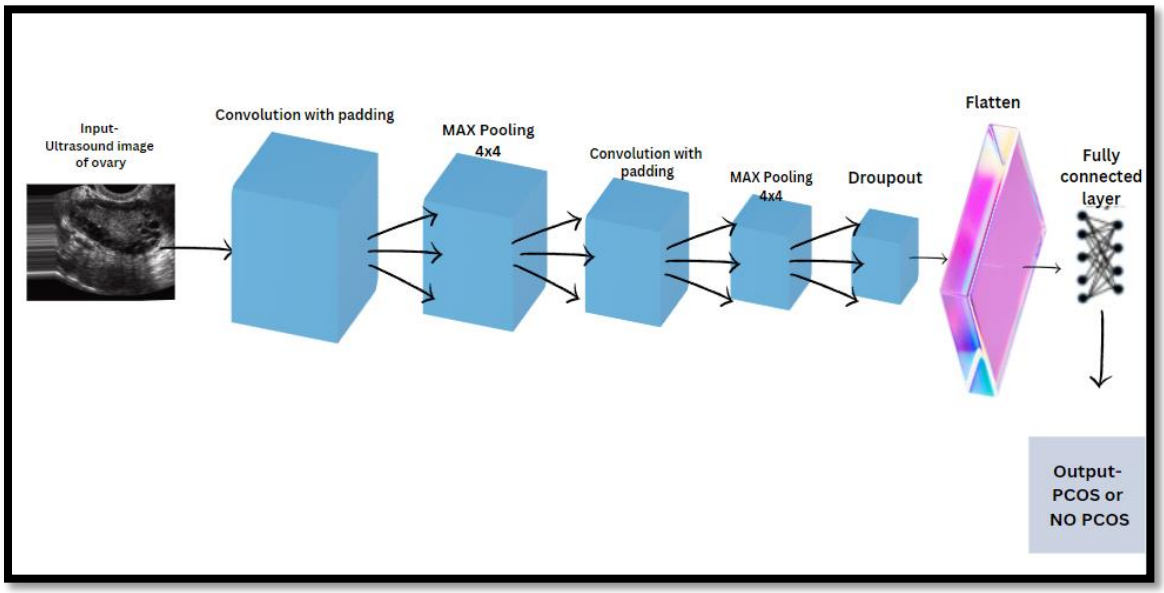


Fig 4.2: Convolutional Neural Network employed in Detection of PCOS using Ultrasound images of ovaries

Following layers have been utilized in this CNN model:

4.3.1. CONVOLUTION LAYER:

It is the fundamental component containing a group of filters or roots with comparatively shorter size as compared to the size of training image [43] In our study, we have kept it 5 x 5. Our Convolution layer has been strengthened via padding that provides extra layer to the external image. In a deep neural network, a weighted sum of input signals is now passed through an activation function, and the output is used as an input. In our study, “reLU” activation has been used which is a Rectified Linear unit function.

Similar convolution is performed after pooling.

4.3.2. POOLING:

With Pooling, overall geographic size of the image is reduced keeping all the important features intact [44] to reduce the computation required. MaxPooling function is unitized in our study which reproduces maximum value for each patch of parameter space via pool size of four by four.

The same kind of pooling has been done after second round of convolution layer.

4.3.3. FLATTEN LAYER:

It basically converges the previous multi-dimentional output into 1-D. The output of this layer creates a one-dimensional array, which helps build the input layer of the classification neural network, where all the components of the array are fed to each neuron. The one-dimensional array created in this step, however, contains an extremely essential & condensed collection of input picture features after conducting all previous CNN procedures. As a result, the feature extraction portion of this machine learning method is believed to run from the "Convolutional layer" to the "Flatten layer."[34]

4.3.4. DENSE LAYER:

The final layer of the CNN architecture, known as the "Dense Layer," is used in this method as the classification layer. This layer functions similar to a feed-forward ANN, and is situated at the end of the CNN framework and, in accordance with the essential design principle of the conventional multilayer perceptron neural system, each neuron in it is interconnected to each and single neuron of the row preceding it [45].

The CNN model was created using the Cross-Entropy function of loss and Adam optimizer of accuracy being the primary assessment parameter for each classifier per epoch. To train the model via images supplied and produce the classification output, machine learning has been carried out in 10 epochs utilising the aforementioned CNN architecture. However, deep learning algorithm analyses the image to take out salient information for classification, negating the need for much image processing prior to training.

```
[88]:
model2 = Sequential()
model2.add(Conv2D(12, (6,6),padding='valid',activation='relu',input_shape=(224,224,3)))
model2.add(MaxPooling2D(pool_size=(6,6)))
model2.add(Conv2D(15, (5,5),padding='valid',activation='relu'))
model2.add(MaxPooling2D(pool_size=(5,5)))
model2.add(Conv2D(10, (3,3),padding='valid',activation='relu'))
# model2.add(Conv2D(256, (5,5),padding='valid',activation='relu'))
model2.add(MaxPooling2D(pool_size=(3,3)))
model2.add(Flatten())
#model2.add(Dense(128,activation='relu'))
#model2.add(Dense(64,activation='relu'))
model2.add(Dense(2,activation='softmax'))
```

Fig 4.3: CNN Models deployed in the study: Convolutional layer, Pooling, Flatten and Dense.

CHAPTER-5

RESULTS:

5.1. ACCURACY OF MODEL

The CNN model that is deployed in this study is a very simple and straightforward algorithm with a short cohort of ultrasound image data. The accuracy came out to be more than 70%.

There are few limitations with the model:

1. Multiple modifications can be done, adding a fine tuning layer where transfer learning can be introduced. This fine tuning layer can use pre-trained models for example- VGGNet16 model having 16 layers, 144 parameters and accuracy of 90% [46].
2. The data size is rather small as compared to other models' training datasets. Collaborations with hospital and diagnosis institutes can help provide much more authentic data than which is available anonymously online.

```
43s 3s/step - loss: 0.6133 - accuracy: 0.6981 - val_loss: 0.5298 - val_accuracy: 0.7865

39s 3s/step - loss: 0.4642 - accuracy: 0.8019 - val_loss: 0.3516 - val_accuracy: 0.8385

38s 3s/step - loss: 0.3377 - accuracy: 0.8494 - val_loss: 0.2544 - val_accuracy: 0.9167
```

	precision	recall	f1-score	support
0	0.70	0.70	0.70	547
1	0.64	0.97	0.77	801
accuracy			0.745	1348
macro avg	0.67	0.835	0.735	1348
weighted avg	0.68	0.87	0.755	1348

Fig 5.1: Overall Model accuracy is found to be 74.5%

5.1.1. PRECISION: Defined by the formula below, also defined as division of true positive (TP) to all the positives output by model. It is the answer to this question- "**When all the predictions were given by our model, how many were actually true?**" [47]

Formula- $$precision = \frac{True\ positives}{True\ positives + false\ positives} \quad (5.1)$$

5.1.2. RECALL: It is our model's True Positive Rate (TPR), It basically answers the following question- "**When model predicted, how many of them were true that should have actually been predicted as true?"** [48]

Formula- $$Recall\ (tpr) = \frac{True\ Positives}{True\ positive + false\ negative} \quad (5.2)$$

5.1.3. F1 SCORE: based on model's precision & model's recall. It measures how efficiently the model presents the trade-off between recall and precision.

Formula- $$F1 - score = 2 * \frac{precison * recall}{precision + recall} \quad (5.3)$$

5.1.4. ACCURACY: Defined as the numeric of optimum predictions in all of the predictions, it is used to measure model's working efficiency [49]

Formula- $$Accuracy = \frac{Number\ of\ correct\ predictions}{size\ of\ data} \quad (5.4)$$

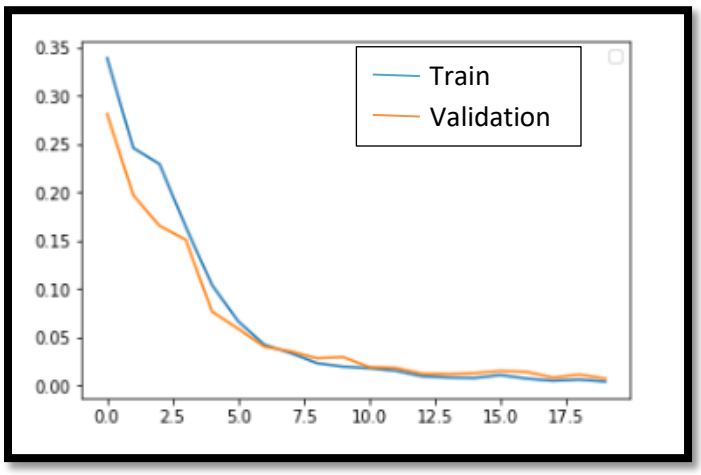


Fig 5.2: With each Epoch, training loss and validation loss is decreasing which explains how well the model is fitting our data, and vice versa, respectively.

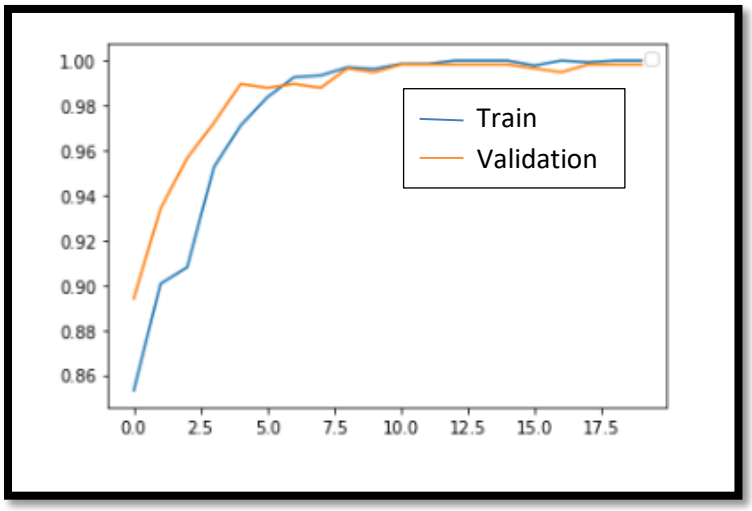


Fig 5.3: With each Epoch, training accuracy and validation accuracy is increasing which explains how well the model is predicting the unseen validation data.

Finally the model is ready to be used and a small demonstration is provided for both infected and not-infected ovary.

5.2. RESULT ANALYSIS

5.2.1. INFECTED Vs. NORMAL OVARY:

The main feature of a polycystic ovary is the diameter and number of follicles present in ovary. If follicles are more than 20, it is said to be a Polycystic ovary[9].

Normal ovary **Infected Ovary**

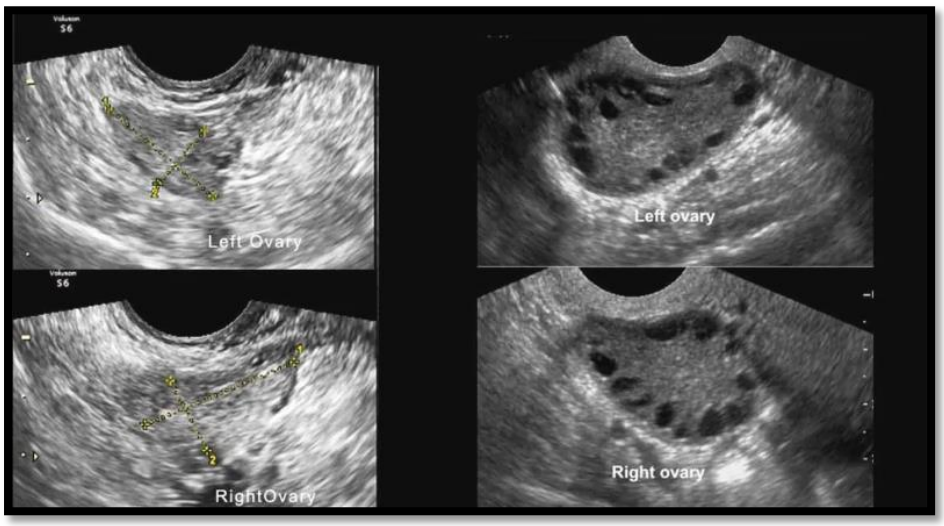


Fig 5.4: Difference in normal and infected ovaries with PCOS as visible in ultrasound images of ovaries.

5.2.2. TESTING MODEL TO PREDICT WITH A TEST ULTRASOUND IMAGE OF AN INFECTED OVARY

```
from keras.preprocessing.image import load_img
image = load_img('../input/pcos-detection-using-ultrasound-images/data/test/infected/img6.jpg', target_size
img = np.array(image)
img = img / 255.0
img = img.reshape(1,224,224,3)
prediction = model.predict(img)
```

Fig 5.5: Input Code to extract test image and check for model's prediction

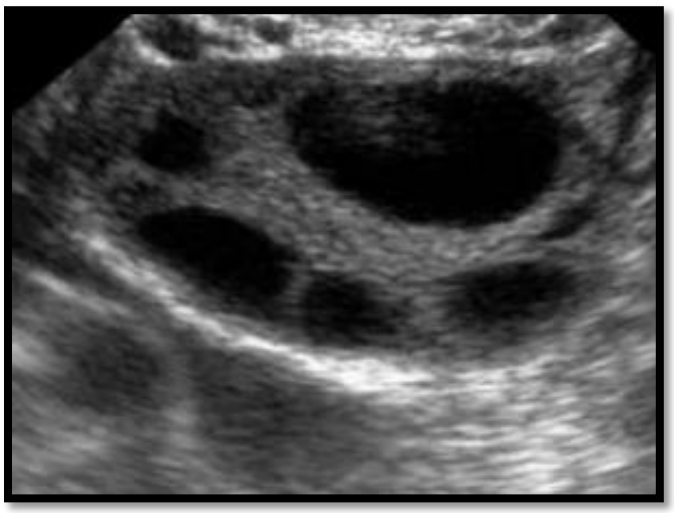


Fig 5.6: Input image: Polycystic ovary

```
[45]:
l={"infected":prediction[0][0],"notinfected":prediction[0][1]}
def get_key(val):
    for key, value in l.items():
         if val == value:
             return key

    return "key doesn't exist"
#label[0][1]

+ Code  + Markdown

j=prediction.max()
get_key(j)

[46]: 'infected'
```

Fig 5.7: Prediction result showing diagnosis as: INFECTED

5.2.3. CONFUSION MATRIX

It is a square matrix used to define performance of a classification algorithm. The design examines the actual objective attributes in relation to the predictions of the model. This provides us with a thorough understanding of the errors that our identification model exhibits [46]. A confusion matrix has 4 variables-

True Positive:

It is a condition when the real and proposed outcome are same and +.

True Negative:

It is a condition when the real and proposed output are same but -.

False Positive:

When the real and predicted output is different, model anticipated it to be a positive class but it actually is negative. It can also be referred to as Type-I error

False Negative:

Condition where the real and proposed outcome is different, model anticipated it to be negative, but actually it is positive. It is also referred to as Type-II error.

We prepared a normalized confusion matrix which treats the sample size in each class as 1, just so to make the analysis consistent [46].

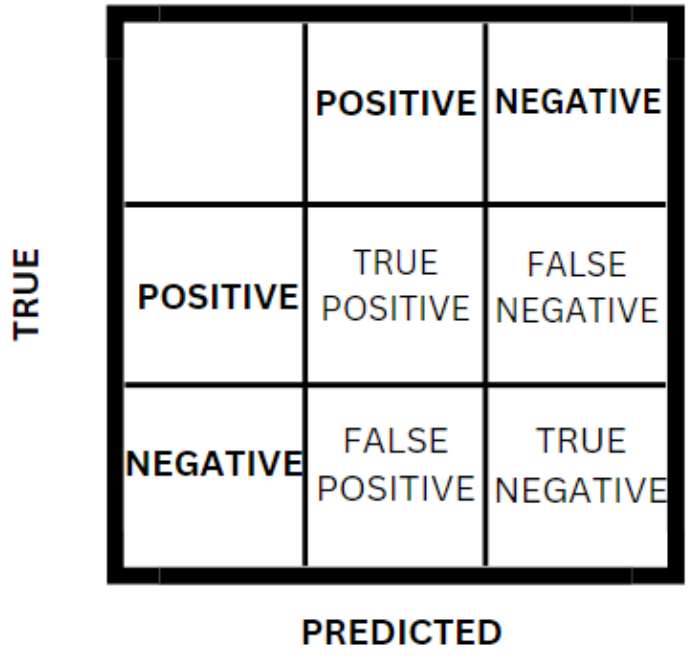


Fig 5.8: A sample confusion matrix

Normalized Confusion matrix-

Columns represent predicted labels whereas rows represent actual labels. It can be inferred that the model works better for diagnosing normal ovaries as compared to infected ovaries

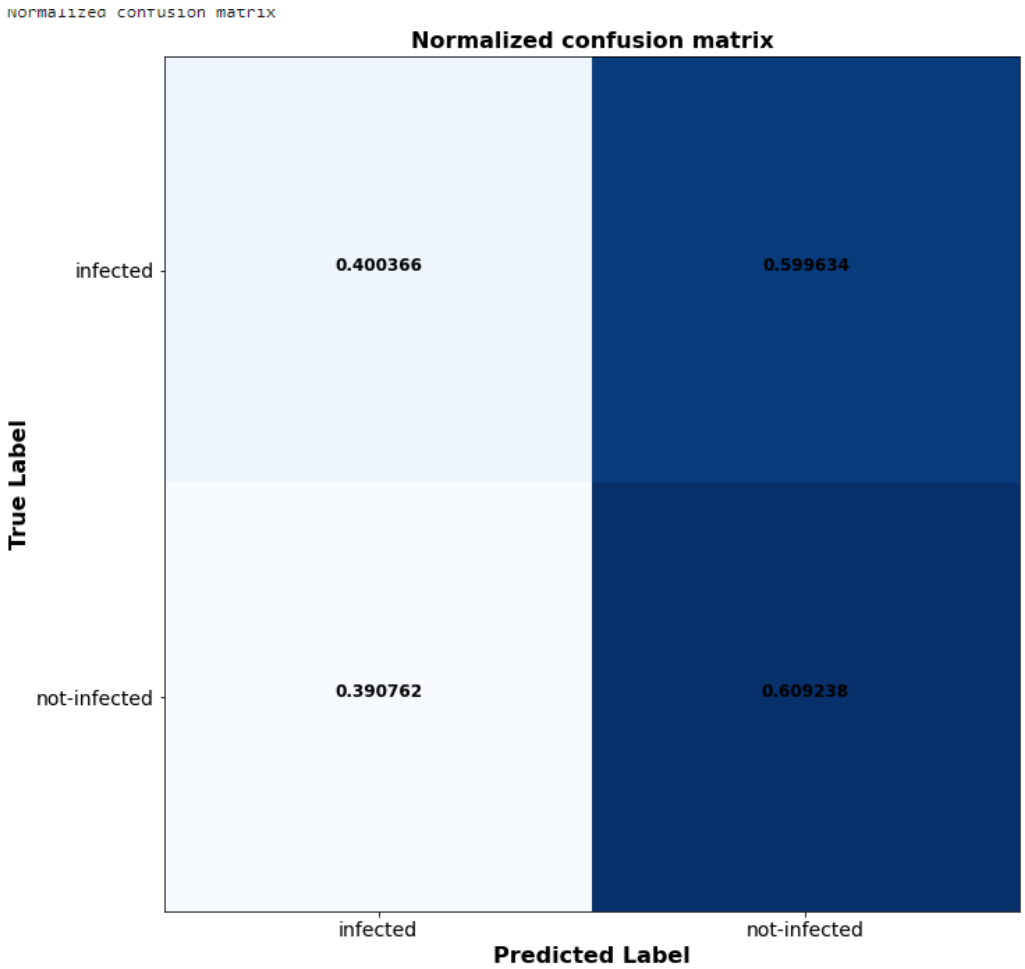


Fig 5.9: Number of true positive (TP) and true negative (TN) are shown by light blue colored boxes whereas False Positive and False negatives are shown by Dark blue colored boxes.

CHAPTER-6

CONCLUSION

Polycystic ovarian syndrome is the most common gynecological disorder affecting reproductive health of millions of women with a chronic timeline. Women can also pass it to their daughters if they are able to conceive otherwise due to anovulatory cycles, women face infertility. Symptoms include hirsuitism, acne, irregular menstrual cycles, weight gain, Insulin resistance aka Type 2 diabetes, etc. The interplay of Hyperandrogenism and Insulin resistance has an important role to play in aetiology of PCOS. Genetic and environmental factors have been reported to accelerate the disease progress. Since the disease is multifactorial and multiple symptoms are visible, it is hard to decode any one absolute treatment plan for it. The treatment hence, is symptomatic. In some, irregular menses is an issue, in others, infertility is the major problem, while some face a death threat from T2D. If we talk about the diagnosis, the situation is not worth the praise. Since there are so many symptoms, it can create a false positive condition; while that symptom would have been because of some other condition, it might lead to a false PCOS diagnosis. The invasive techniques are not always patient compliant, and the ultrasound images might be misread even by professionals sometimes.

Artificial intelligence becomes a helping hand in such cases and takes into account all kinds of features that we want to diagnose in case of PCOS. Since it is a computed algorithm, it is going to work exactly same and correct every time until the data provided to it is appropriate. Artificial intelligence is the modern computer science technique through which the automation and human labor is reduced to a very large extent. There are multiple sub-disciplines in Artificial intelligence, like Machine learning and deep learning. Although both of them are almost similar and interconnected, deep learning uses principles of machine learning to prepare advanced solutions to advanced human problems. DL & ML have proven its worth in almost all of the sectors, and in healthcare too. Research including ML &DL is increasing day by day with more opportunities opening since the increase of digitized data available online.

Omics approach is one of the recent advancements in biotechnology, and has opened multiple ways of looking at and working in biological science. AI technology makes it feasible to automate diagnostics, patient screening, pain anticipating, drug development, treatments including computerised drug delivery, precision healthcare, future outcomes,

and decision-making assistance [50]. Machine learning has applications in drug discovery [51]. Molecular docking of drugs is a very useful approach for drug repurposing and machine learning plays a crucial role in validating the results [51] In cases of cancers, neurodegenerative disorders, Machine learning has become a tool in there prediction at early treatable stages. Self-diagnostic models using non-invasive parameters like weight, acne, hirsuitism, etc., and invasive parameters like AMH concentration and number of follicles in ovaries, are also available for PCOS which are built on Machine learning [52]

In this study, we tried to build a simple and straightforward Deep learning model to predict Polycystic Ovarian Syndrome using a large dataset of ultrasound images of infected and normal ovaries available online. The model employed in the present study was Convolutional Neural Network that works in layers and transfers information to next layers. With an accuracy of 74.5%, the model was working fine and can be improved. Comparative analysis parameters like precision, recall and FI-Score were also determined. A normalized confusion matrix was prepared to null out the confusion model is facing.

With more datasets and increase in number of ML models in our CNN model for prediction of PCOS, accuracy can easily be improved. Following are the limitations that we faced:

1. Multiple modifications can be done, adding a fine tuning layer where transfer learning can be introduced. This fine tuning layer can use pre-trained models for example- VGGNet16 model having 16 layers, 144 parameters and accuracy of 90%. [46]
2. The data size is rather small as compared to other models' training datasets. Collaborations with hospital and diagnosis institutes can help provide much more authentic data than which is available anonymously online.

CHAPTER – 7

FUTURE PROSPECTS OF ARTIFICIAL INTELLIGENCE IN PCOS

It is believed that a successful artificial intelligence technique to identify PCOS from ovarian ultrasound pictures could be a useful way to restore thousands of women's reproductive health. Recent research in detection of PCOS using machine learning has experienced a sudden boom. For example, studies done by [53][54][55] used diverse image processing methods in follicle detection of ovaries, then employed traditional or CNN models to diagnose PCOS in clinical samples. Due to lack of data which is authentic, easy to process and used in algorithm, these Convolutional neural networks lag behind in terms of accuracy. [6] study showed that with an AUC value of 0.978, their model was successfully able to detect PCOS, and promises a huge potential of Deep learning in detection of PCOS and other disease alike.

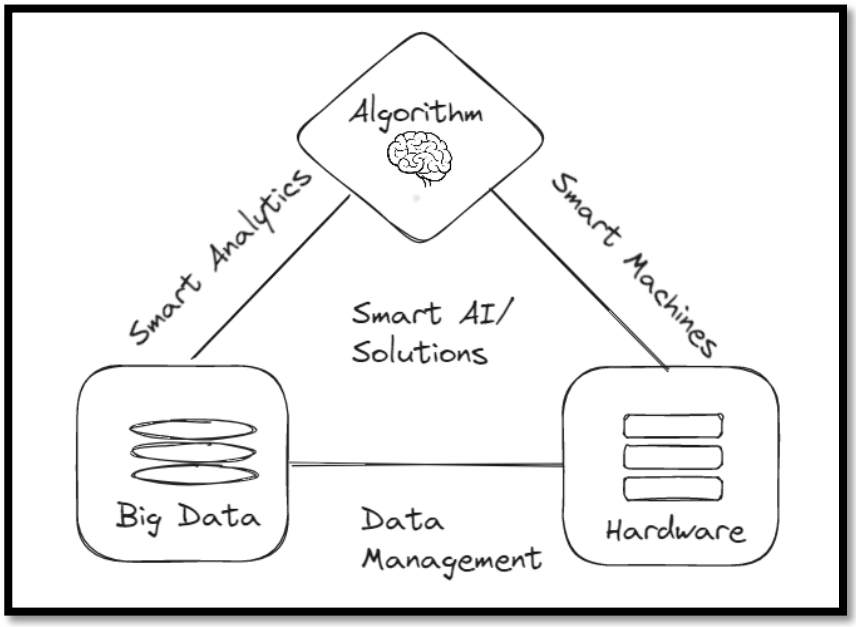


Fig 7.1: The future of Artificial Intelligence will require a strong interrelation of each of the elements: Big Data, Hardware and Algorithm to produce Smart solutions

AI technology makes it feasible to automate diagnostics, patient screening, pain anticipating, drug development, treatments including computerised drug delivery, precision healthcare, future outcomes, and decision-making assistance [50]

Explainable AI or XAI is a validation tool for AI or ML models being used in providing a clinical validation for in-silico research. If we accept this study, it can be said that the

laborious clinical validations can be ruled out with the help of Artificial intelligence. As proposed in [50], future work can be based on constructing a personalized application for all of the potential patient population, that is capable of performing PCOS screening in real time without any invasive technique. This can provide a risk assessment analysis to the users, and warn them if they show any symptoms beforehand [56].

This work requires more database which is exhaustive and top notch quality as the application is directly catering to the customers. Any misdiagnosis or faulty analysis can risk a life. This will help associating healthcare workers and informaticians keeping in mind that the model is as easy to interpret to the healthcare professionals as it is to the computer science expert.

REFERENCES:

[1] T. H., D. A., and M. L., "Polycystic ovary syndrome: A complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan," *BMC Med.*, vol. 8, no. 41, pp. 1–10, 2010, [Online]. Available: http://www.biomedcentral.com/1741-7015/8/41%0Ahttp://www.biomedcentral.com/1741-7015/8/41%5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed9&NEWS=N&AN=2010416064

[2] T. M. Barber and S. Franks, "The link between polycystic ovary syndrome and both Type 1 and Type 2 diabetes mellitus: What do we know today?," *Women's Heal.*, vol. 8, no. 2, pp. 147–154, 2012, doi: 10.2217/whe.11.94.

[3] P. Christopoulos *et al.*, "Genetic variants in TCF7L2 and KCNJ11 genes in a Greek population with polycystic ovary syndrome," *Gynecol. Endocrinol.*, vol. 24, no. 9, pp. 486–490, 2008, doi: 10.1080/09513590802196379.

[4] J. M. C. Frary, K. P. Bjerre, D. Glintborg, and P. Ravn, "The effect of dietary carbohydrates in women with polycystic ovary syndrome: A systematic review," *Minerva Endocrinol.*, vol. 41, no. 1, pp. 57–69, 2016.

[5] N. Bakkar *et al.*, "Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis," *Acta Neuropathol.*, vol. 135, no. 2, pp. 227–247, 2018, doi: 10.1007/s00401-017-1785-8.

[6] A. H. Hande, M. S. Chaudhary, A. R. Gadbail, P. R. Zade, M. N. Gawande, and S. K. Patil, "Role of hypoxia in malignant transformation of oral submucous fibrosis," *J. Datta Meghe Inst. Med. Sci. Univ.*, vol. 13, no. 1, pp. 38–43, 2018, doi: 10.4103/jdmimsu.jdmimsu.

[7] M. Sumathi, P. Chitra, R. Sakthi Prabha, and K. Srilatha, "Study and detection of PCOS related diseases using CNN," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1070, p. 012062, 2021, doi: 10.1088/1757-899x/1070/1/012062.

[8] A. K. M. Salman Hosain, M. H. K. Mehedi, and I. E. Kabir, "PCONet: A Convolutional Neural Network Architecture to Detect Polycystic Ovary Syndrome (PCOS) from Ovarian Ultrasound Images," *8th Int. Conf. Eng. Emerg. Technol. ICEET 2022*, no. October, pp. 27–28, 2022, doi: 10.1109/ICEET56468.2022.10007353.

[9] S. Kumari, "Classification of PCOS / PCOD Using Transfer Learning and GAN

Architectures to Generate Pseudo Ultrasound Images," 2021.

[10] Y. V. Louwers and J. S. E. Laven, "Characteristics of polycystic ovary syndrome throughout life," *Ther. Adv. Reprod. Heal.*, vol. 14, p. 263349412091103, 2020, doi: 10.1177/2633494120911038.

[11] Z. Dehghan, S. Mohammadi-Yeganeh, M. Sameni, S. A. Mirmotalebisohi, H. Zali, and M. Salehi, "Repurposing new drug candidates and identifying crucial molecules underlying PCOS Pathogenesis Based On Bioinformatics Analysis," *DARU, J. Pharm. Sci.*, vol. 29, no. 2, pp. 353–366, 2021, doi: 10.1007/s40199-021-00413-9.

[12] M. R. Jones *et al.*, "Systems Genetics Reveals the Functional Context of PCOS Loci and Identifies Genetic and Molecular Mechanisms of Disease Heterogeneity," *PLoS Genet.*, vol. 11, no. 8, pp. 1–17, 2015, doi: 10.1371/journal.pgen.1005455.

[13] A. E. Sørensen, P. B. Udesen, M. L. Wissing, A. L. M. Englund, and L. T. Dalgaard, "MicroRNAs related to androgen metabolism and polycystic ovary syndrome," *Chem. Biol. Interact.*, vol. 259, pp. 8–16, 2016, doi: 10.1016/j.cbi.2016.06.008.

[14] L. E. Nicol, T. D. O'Brien, D. A. Dumesic, T. Grogan, A. F. Tarantal, and D. H. Abbott, "Abnormal infant islet morphology precedes insulin resistance in PCOS-like monkeys," *PLoS One*, vol. 9, no. 9, pp. 1–8, 2014, doi: 10.1371/journal.pone.0106527.

[15] M. Gholinezhad, M. Gholsorkhtabaramiri, S. Esmaeilzadeh, and A. Ghanbarpour, "Insulin resistance and adverse metabolic profile in overweight/obese and normal weight of young women with polycystic ovary syndrome," *Casp. J. Intern. Med.*, vol. 9, no. 3, pp. 260–267, 2018, doi: 10.22088/cjim.9.3.260.

[16] A. S. Laganà *et al.*, "Metabolism and Ovarian Function in PCOS Women: A Therapeutic Approach with Inositols," *Int. J. Endocrinol.*, vol. 2016, no. Table 1, 2016, doi: 10.1155/2016/6306410.

[17] Q. Gao *et al.*, "Exploration of molecular features of PCOS with different androgen levels and immune-related prognostic biomarkers associated with implantation failure," *Front. Endocrinol. (Lausanne).*, vol. 13, no. August, pp. 1–17, 2022, doi: 10.3389/fendo.2022.946504.

[18] H. Ding *et al.*, "Resistance to the Insulin and Elevated Level of Androgen: A Major Cause of Polycystic Ovary Syndrome," *Front. Endocrinol. (Lausanne).*,

vol. 12, no. October, pp. 1–14, 2021, doi: 10.3389/fendo.2021.741764.

[19] D. Fahs, D. Salloum, M. Nasrallah, and G. Ghazeeri, “Polycystic Ovary Syndrome: Pathophysiology and Controversies in Diagnosis,” *Diagnostics*, vol. 13, no. 9, pp. 1–13, 2023, doi: 10.3390/diagnostics13091559.

[20] A. E. Taylor *et al.*, “Determinants of abnormal gonadotropin secretion in clinically defined women with polycystic ovary syndrome,” *J. Clin. Endocrinol. Metab.*, vol. 82, no. 7, pp. 2248–2256, 1997, doi: 10.1210/jc.82.7.2248.

[21] D. T. Baird, C. S. Corker, D. W. Davidson, W. M. Hunter, E. A. Michie, and P. F. A. Vanlook, “Pituitary-ovarian relationships in polycystic ovary syndrome,” *J. Clin. Endocrinol. Metab.*, vol. 45, no. 4, pp. 798–809, 1977, doi: 10.1210/jcem-45-4-798.

[22] M. K. Murphy, J. E. Hall, J. M. Adams, H. Lee, and C. K. Welt, “Polycystic ovarian morphology in normal women does not predict the development of polycystic ovary syndrome,” *J. Clin. Endocrinol. Metab.*, vol. 91, no. 10, pp. 3878–3884, 2006, doi: 10.1210/jc.2006-1085.

[23] M. A. Haendel, C. G. Chute, and P. N. Robinson, “Classification, Ontology, and Precision Medicine,” *N. Engl. J. Med.*, vol. 379, no. 15, pp. 1452–1462, 2018, doi: 10.1056/nejmra1615014.

[24] P. D. Kilmer, “Review Article: Review Article,” *Journalism*, vol. 11, no. 3, pp. 369–373, 2010, doi: 10.1177/1461444810365020.

[25] B. C. J. M. Fauser, “Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome,” *Fertil. Steril.*, vol. 81, no. 1, pp. 19–25, 2004, doi: 10.1016/j.fertnstert.2003.10.004.

[26] A. Dokras, S. Saini, M. Gibson-Helm, J. Schulkin, L. Cooney, and H. Teede, “Gaps in knowledge among physicians regarding diagnostic criteria and management of polycystic ovary syndrome,” *Fertil. Steril.*, vol. 107, no. 6, pp. 1380-1386.e1, 2017, doi: 10.1016/j.fertnstert.2017.04.011.

[27] H. J. Teede *et al.*, “Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome,” *Hum. Reprod.*, vol. 33, no. 9, pp. 1602–1618, 2018, doi: 10.1093/humrep/dey256.

[28] J. A. Lundgren, S. H. Kim, C. M. Burt Solorzano, C. R. McCartney, and J. C. Marshall, “Progesterone Suppression of Luteinizing Hormone Pulse Frequency in Adolescent Girls With Hyperandrogenism: Effects of Metformin,” *J. Clin. Endocrinol. Metab.*, vol. 103, no. 1, pp. 263–270, 2018, doi: 10.1210/jc.2017-

02068.

[29] M. Lu *et al.*, “CREBZF regulates testosterone production in mouse Leydig cells,” *J. Cell. Physiol.*, vol. 234, no. 12, pp. 22819–22832, 2019, doi: 10.1002/jcp.28846.

[30] F. Qu *et al.*, “A molecular mechanism underlying ovarian dysfunction of polycystic ovary syndrome: Hyperandrogenism induces epigenetic alterations in the granulosa cells,” *J. Mol. Med.*, vol. 90, no. 8, pp. 911–923, 2012, doi: 10.1007/s00109-012-0881-4.

[31] A. J. T. Pedersen, T. B. Stage, D. Glintborg, M. Andersen, and M. M. H. Christensen, “The Pharmacogenetics of Metformin in Women with Polycystic Ovary Syndrome: A Randomized Trial,” *Basic Clin. Pharmacol. Toxicol.*, vol. 122, no. 2, pp. 239–244, 2018, doi: 10.1111/bcpt.12874.

[32] G. Rena, D. G. Hardie, and E. R. Pearson, “The mechanisms of action of metformin,” *Diabetologia*, vol. 60, no. 9, pp. 1577–1585, 2017, doi: 10.1007/s00125-017-4342-z.

[33] A. H. Balen *et al.*, “The management of anovulatory infertility in women with polycystic ovary syndrome: An analysis of the evidence to support the development of global WHO guidance,” *Hum. Reprod. Update*, vol. 22, no. 6, pp. 687–708, 2016, doi: 10.1093/humupd/dmw025.

[34] T. Wang *et al.*, “A review on medical imaging synthesis using deep learning and its clinical applications,” *J. Appl. Clin. Med. Phys.*, vol. 22, no. 1, pp. 11–36, 2021, doi: 10.1002/acm2.13121.

[35] F. Sohil, M. U. Sohali, and J. Shabbir, “An introduction to statistical learning with applications in R,” *Stat. Theory Relat. Fields*, vol. 6, no. 1, pp. 87–87, 2022, doi: 10.1080/24754269.2021.1980261.

[36] T. et. all. Hastie, “Statistics The Elements of Statistical Learning,” *Springer Ser. Stat.*, vol. 27, no. 2, p. 745, 2009, [Online]. Available: http://www.springerlink.com/index/D7X7KX6772HQ2135.pdf

[37] V. Thakre, “PCOcare: PCOS Detection and Prediction using Machine Learning Algorithms,” *Biosci. Biotechnol. Res. Commun.*, vol. 13, no. 14, pp. 240–244, 2020, doi: 10.21786/bbrc/13.14/56.

[38] Z. Chen, T. Anglea, and Y. Zhang, “Optimal synchronization in pulse-coupled oscillator,” *PNAS Nexus*, vol. 2, no. 4, pp. 1–7, 2023, doi: 10.1093/pnasnexus/pgad102.

[39] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. Peter Campbell, “Introduction to machine learning, neural networks, and deep learning,” *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, pp. 1–12, 2020, doi: 10.1167/tvst.9.2.14.

[40] T. F. Gonzalez, “Handbook of approximation algorithms and metaheuristics,” *Handb. Approx. Algorithms Metaheuristics*, pp. 1–1432, 2007, doi: 10.1201/9781420010749.

[41] W. Lv *et al.*, “Deep Learning Algorithm for Automated Detection of Polycystic Ovary Syndrome Using Scleral Images,” *Front. Endocrinol. (Lausanne).*, vol. 12, no. January, pp. 1–8, 2022, doi: 10.3389/fendo.2021.789878.

[42] V. Gulshan *et al.*, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *JAMA - J. Am. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, 2016, doi: 10.1001/jama.2016.17216.

[43] W. Yin, S. Mostafa, and F. X. Wu, “Diagnosis of Autism Spectrum Disorder Based on Functional Brain Networks with Deep Learning,” *J. Comput. Biol.*, vol. 28, no. 2, pp. 146–165, 2021, doi: 10.1089/cmb.2020.0252.

[44] A. Patil and M. Rane, “Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition,” *Smart Innov. Syst. Technol.*, vol. 195, pp. 21–30, 2021, doi: 10.1007/978-981-15-7078-0_3.

[45] S. A. Suha and M. N. Islam, “An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–16, 2022, doi: 10.1038/s41598-022-21724-0.

[46] A. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran, “A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning,” *Circuits, Syst. Signal Process.*, vol. 39, no. 2, pp. 757–775, 2020, doi: 10.1007/s00034-019-01246-3.

[47] W. Liu, S. Wang, Z. Ye, P. Xu, X. Xia, and M. Guo, “Prediction of lung metastases in thyroid cancer using machine learning based on SEER database,” *Cancer Med.*, vol. 11, no. 12, pp. 2503–2515, 2022, doi: 10.1002/cam4.4617.

[48] G. H. Fu, L. Z. Yi, and J. Pan, “Tuning model parameters in class-imbalanced learning with precision-recall curve,” *Biometrical J.*, vol. 61, no. 3, pp. 652–664, 2019, doi: 10.1002/bimj.201800148.

[49] D. T. Anh, H. Takakura, M. Asai, N. Ueda, and H. Shojaku, “Application of

machine learning in the diagnosis of vestibular disease," *Sci. Rep.*, vol. 12, no. 1, pp. 1–10, 2022, doi: 10.1038/s41598-022-24979-9.

[50] V. V. Khanna, K. Chadaga, N. Sampathila, S. Prabhu, V. Bhandage, and G. K. Hegde, "A Distinctive Explainable Machine Learning Framework for Detection of Polycystic Ovary Syndrome," *Appl. Syst. Innov.*, vol. 6, no. 2, p. 32, 2023, doi: 10.3390/asi6020032.

[51] R. Gupta, D. Srivastava, M. Sahu, S. Tiwari, R. K. Ambasta, and P. Kumar, *Artificial intelligence to deep learning: machine intelligence approach for drug discovery*, vol. 25, no. 3. Springer International Publishing, 2021. doi: 10.1007/s11030-021-10217-3.

[52] A. Zigarelli, Z. Jia, and H. Lee, "Machine-Aided Self-diagnostic Prediction Models for Polycystic Ovary Syndrome: Observational Study," *JMIR Form. Res.*, vol. 6, no. 3, 2022, doi: 10.2196/29967.

[53] A. Mandal, D. Saha, and M. Sarkar, *Follicle Segmentation Using K-Means Clustering from Ultrasound Image of Ovary*, vol. 1255, no. February 2021. Springer Singapore, 2021. doi: 10.1007/978-981-15-7834-2_51.

[54] P. G. YILMAZ and G. ÖZMEN, "Follicle Detection for Polycystic Ovary Syndrome by using Image Processing Methods," *Int. J. Appl. Math. Electron. Comput.*, vol. 8, no. 4, pp. 203–208, 2020, doi: 10.18100/ijamec.803400.

[55] C. Gopalakrishnan and M. Iyapparaja, "Active contour with modified Otsu method for automatic detection of polycystic ovary syndrome from ultrasound image of ovary," *Multimed. Tools Appl.*, vol. 79, no. 23–24, pp. 17169–17192, 2020, doi: 10.1007/s11042-019-07762-3.

[56] S. Alam Suha and M. N. Islam, "Exploring the Dominant Features and Data-driven Detection of Polycystic Ovary Syndrome through Modified Stacking Ensemble Machine Learning Technique," *Heliyon*, vol. 9, no. 3, p. e14518, 2023, doi: 10.1016/j.heliyon.2023.e14518.

APPENDIX

A.1. Modules used in preparing CNN Code:

1. **TensorFlow**: It is framework known for its libraries that are free to use/ open source aiding in tricky numeric computations. Developed by Google, available in C++ and python languages, it is one of the most used frameworks in Machine learning algortihms.
2. Keras: It acts as the interface for TF library to work on artificial neural network models. It is user-friendly, fast and modular and built specially for deep learning algorithms to speed up the process.
3. Categorical Cross entropy: It is used as a loss function for models with 2 or more than 2 output; multiclass classification. It is imported from keras.losses and the output is given a value from 0 to 1.

A.2 Code for using categorical cross entropy to calculate losses in model.

```
In [12]:
from tensorflow.keras.losses import CategoricalCrossentropy
model1.compile(
  optimizer='adam',
  loss=CategoricalCrossentropy(),
  metrics=['accuracy'])
```

A.3 CNN Model using sequential() function showing trainable parameters

```
In [11]: model1.summary()

Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 220, 220, 10)      760
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 55, 55, 10)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 51, 51, 12)        3012
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 12, 12, 12)        0
_________________________________________________________________
flatten (Flatten)            (None, 1728)              0
_________________________________________________________________
dense (Dense)                (None, 2)                 3458
=================================================================
Total params: 7,230
Trainable params: 7,230
Non-trainable params: 0
_________________________________________________________________
```

A.4 Code for preparing Confusion Matrix

```
import matplotlib.pyplot as plt
import itertools
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
def plot_confusion_matrix(cm, classes,
                          normalize= False,
                          title='Confusion matrix',
                          cmap=plt.cm.Blues):
    if normalize:
        cm= cm.astype('float') / cm.sum(axis=1)[:,np.newaxis]
        print("Normalized confusion matrix")
    else:
        print('Confusion matrix, without normalization')

    plt.imshow(cm, interpolation = 'nearest', cmap = cmap)
    plt.title(title, weight = 'bold', fontsize= 16)
    tick_marks = np.arange(len(classes))
    plt.xticks(tick_marks, classes, fontsize= 14)
    plt.yticks(tick_marks, classes, fontsize = 14)

    fmt = '2f' if normalize else 'd'
    thresh = cm.max()/2
    for i, j in itertools.product(range(cm.shape[0]), range(cm.shape[1])):
        plt.text(j,i, format(cm[i,j], fmt),
                 horizontalalignment = "center", fontsize = 12, weight = 'bold',
                 color = "black" if cm[i,j] > thresh else "white")
    plt.tight_layout()
    plt.ylabel('True Label', fontsize = 16, weight = 'bold')
    plt.xlabel('Predicted Label', fontsize = 16, weight = 'bold')

#compute consufion matrix
cnf_matrix = confusion_matrix(train_it.classes, y_pred)
np.set_printoptions(precision = 2)

#Plot non-normalized confusion matrix
plt.figure(figsize=(10,10))
plot_confusion_matrix(cnf_matrix, classes= ['infected', 'not-infected'], normalize= True, title='Normalized confusion matrix'
plt.show()
```

PROOF OF CONFERENCE

2023 8th International conference for Convergence in Technology (I2CT)

Acceptance e-mail:

---------- Forwarded message ---------
From: **Microsoft CMT** <email@msr-cmt.org>
Date: Sun, 12 Mar 2023, 19:39
Subject: ACCEPTANCE NOTIFICATION - 8th I2CT 2023
To: Yasha Hasija <yashahasija06@gmail.com>

Dear Yasha Hasija

Paper ID / Submission ID : 2326

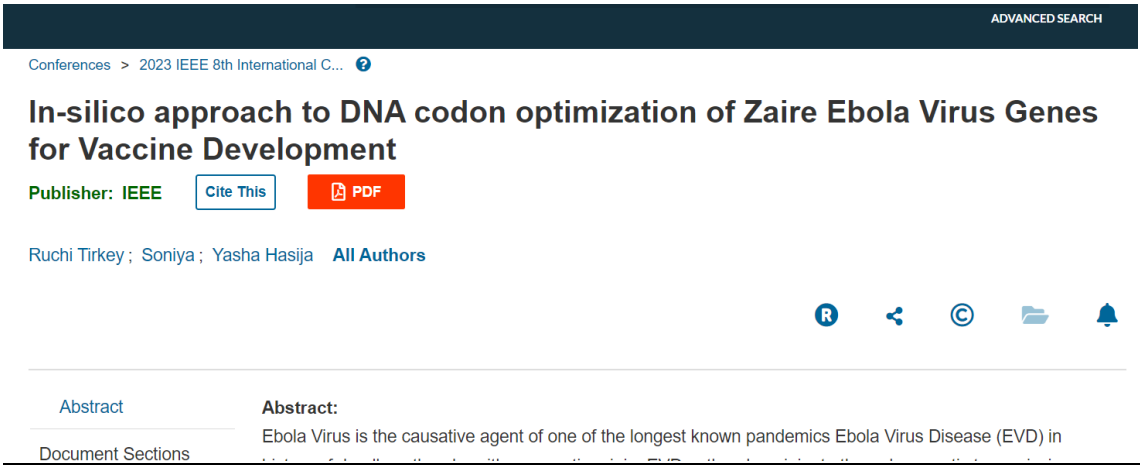
Title : In-silico approach to DNA codon optimization of Zaire Ebola virus genes for vaccine development

Greeting from I2CT - The premier conference for the latest discoveries in Convergence in Technology in Asia pacific.

We are pleased to inform you that your paper has been accepted for the Oral Presentation as a full paper for the- "IEEE 2023 8th International Conference for Convergence in Technology (I2CT),Pune, Maharashtra, India with following reviewers' comment.

All accepted and presented papers will be submitted to IEEE Xplore for the further publication and will be indexed by Ei Compendex and Scopus Indexing.

Published on 23 May 2023 with DOI: 10.1109/I2CT57861.2023.10126407



CERTIFICATE

8th International Conference for Convergence in Technology (I2CT)

IEEE BOMBAY SECTION

7th – 9th April 2023

Certificate

This is to certify that Dr./Prof./Mr./Ms. ***Soniya*** *has presented paper entitled* ***In-silico approach to DNA codon optimization of Zaire Ebola virus genes for vaccine development*** *in 8th International Conference for Convergence in Technology (I2CT) during 7th & 9th April 2023.*

Dr. Chanakya Kumar Jha
General Chair - I2CT

turnitin

Similarity Report ID: oid:27535:36360714

PAPER NAME

masters thesis (2).docx

WORD COUNT	CHARACTER COUNT
10885 Words	**59774 Characters**
PAGE COUNT	FILE SIZE
54 Pages	**2.3MB**
SUBMISSION DATE	REPORT DATE
May 28, 2023 8:25 PM GMT+5:30	**May 28, 2023 8:26 PM GMT+5:30**

9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 6% Internet database
- Crossref database
- 7% Submitted Works database
- 4% Publications database
- Crossref Posted Content database

Excluded from Similarity Report

- Bibliographic material

turnitin

Similarity Report ID: oid:27535:36360714

9% Overall Similarity

Top sources found in the following databases:

- 6% Internet database
- Crossref database
- 7% Submitted Works database
- 4% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	dspace.dtu.ac.in:8080 Internet	4%
2	University of Southampton on 2022-05-05 Submitted works	<1%
3	Liverpool John Moores University on 2022-03-08 Submitted works	<1%
4	researchgate.net Internet	<1%
5	Ruchi Tirkey, Soniya, Yasha Hasija. "In-silico approach to DNA codon o... Crossref	<1%
6	downloads.hindawi.com Internet	<1%
7	nature.com Internet	<1%
8	tandfonline.com Internet	<1%

9	Michigan Technological University on 2023-04-26 Submitted works	<1%
10	Rajiv Gandhi Indian Institute of Management on 2023-03-01 Submitted works	<1%
11	"Polycystic Ovary Syndrome", Springer Science and Business Media LL... Crossref	<1%
12	Glasgow Caledonian University on 2021-05-07 Submitted works	<1%
13	tcmjmed.com Internet	<1%
14	Liverpool John Moores University on 2022-12-24 Submitted works	<1%
15	mdpi.com Internet	<1%
16	Shuxia Li, Dongyi Zhu, Hongmei Duan, Qihua Tan. "Genetic investigatio... Crossref	<1%
17	West Coast University on 2021-06-28 Submitted works	<1%
18	dr.ntu.edu.sg Internet	<1%
19	eprints.kfupm.edu.sa Internet	<1%
20	Endeavour College of Natural Health on 2012-05-13 Submitted works	<1%

21	University of Auckland on 2022-10-13 Submitted works	<1%
22	Université Saint-Esprit Kaslik on 2021-05-03 Submitted works	<1%
23	West Coast University on 2022-03-21 Submitted works	<1%
24	Higher Education Commission Pakistan on 2015-12-17 Submitted works	<1%
25	Kaplan Professional on 2023-04-27 Submitted works	<1%
26	University of Hertfordshire on 2023-04-19 Submitted works	<1%
27	ebin.pub Internet	<1%
28	gssrr.org Internet	<1%
29	Cardiff University on 2019-05-06 Submitted works	<1%
30	MGH Institute of Health Professions on 2020-04-08 Submitted works	<1%
31	Roland, Alison V., and Suzanne M. Moenter. "Reproductive neuroendoc... Crossref	<1%
32	University of Worcester on 2021-01-04 Submitted works	<1%

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

CANDIDATE'S DECLARATION

I Soniya, Roll Number: 2K21/MSCBIO/50, student of M.Sc. Biotechnology, hereby declare that the work which is presented in the Major Project entitled — **A Study of paradigm shift in diagnosing diseases using Artificial Intelligence: Case study on Polycystic Ovarian Syndrome** in partial fulfillment of the requirement for the award of the degree of Master of Science in Biotechnology and submitted to the Department of Biotechnology, Delhi Technological University, Delhi, is an authentic record of my own carried out work during the period from January-May 2023, under the supervision of Prof. Yasha Hasija. The matter presented in this report has not been submitted by me for the award for any other degree of this or any other Institute/University.

The following attached work has been accepted in IEEE conference with the following details:

Title of the Paper: In-silico approach to DNA codon optimization of Zaire Ebola virus genes for vaccine development

Author Names: Soniya, Tirkey, Ruchi and Hasija, Yasha

Name of Conference: The International Conference for Convergence in Technology (I2CT)

Conference Date and Venue: 7th-9th April 2023/ Venue- Hotel Vivanta Pune, Hinjawadi, Maharshtra, India

Registration: Done

Status of Paper (Accepted/Communicated/Published): Published

Date of Paper Communication: 28 February, 2023

Date of Paper Acceptance: 12 March, 2023

Date of Paper Publication: 23 May, 2023

Soniya

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi - 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "**A Study of paradigm shift in diagnosing diseases using Artificial Intelligence: Case study on Polycystic Ovarian Syndrome**" which is submitted by **Soniya (2K21/MSCBIO/50)**, Department of Biotechnology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Science is recorded for the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any degree or any diploma to this university or elsewhere.

Place: Delhi
Date: 30/05/2023.

Yashasi 30.05.23

Prof. Yasha Hasija
(Supervisor)
Professor
Department of Biotechnology
Delhi Technological University

30/05/2023

Prof. Pravir Kumar
Head of Department

Department of Biotechnology
Delhi Technological University

ACKNOWLEDGEMENT

This is my heartiest gratitude to my supervisor Dr. Yasha Hasija, who has never failed to provide me all the best opportunities and resources for the successful completion of this dissertation. This research has been done in the field of my utmost interest, to which she has been a guiding light throughout the period of six months. A source of inspiration, Dr. Yasha Hasija, has showered me with wisdom and dynamism.

Sincere gratitude to Dr. Pravir Kumar, Head of Department of Biotechnology, DTU, who provided all the necessary resources required for the completion of this project work.

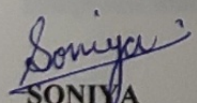
Big thanks to my research mate, Ruchi Tirkey, along with whom I share best of my project memories. It was her, who motivated me in tough times and we were able to submit conference paper in IEEE.

Special thanks to Mrs Jayshree Meena, Miss Neha Kumari for bein the wonderful mentors and helping me with all the knowledge they have.

The lab staff of department of biotechnology has always had my back, and I am blessed to have their constant support and help whenever I needed it.

All my classmates, including Bharmjeet Dahiya, Manju Dahiya, Tanya Srivastava, and lab mates of Computational Biology lab have been nothing less than a blessing who have helped me whenever I stuck.

Last but not least, I acknowledge the support, care and love given by my parents and family. Without them, I won't be here.



SONIYA
2K21/MSCBIO/50