A DISSERTATION on
RETINAL VESSEL DETECTION USING RESIDUAL Y-NET

Submitted in partial fulfillment of the requirement for the award of the degree of
MASTER IN TECHNOLOGY
IN
(INFORMATION SYSTEMS)
Submitted by

ANINDITA ROY
2k19/ISY/01

Under the supervision of

PROF. KAPIL SHARMA

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

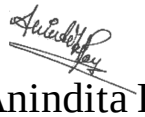
JUNE, 2021

CANDIDATE's DECLARATION

I hereby declare that the work presented in this dissertation/thesis entitled "RETINAL VESSEL DETECTION USING RESIDUAL Y-NET", in partial fulfillment of the requirements for the award of the MASTER OF TECHNOLOGY degree in Information Systems submitted in Information Technology Department at DELHI TECHNOLOGICAL UNIVERSITY, New Delhi, is an authentic record of my own work carried out during my degree under the guidance of Prof. Kapil Sharma.

Date: June, 2021

Place: New Delhi



Anindita Roy (2K19/ISY/01)

CERTIFICATE

This is to certify that Anindita Roy (2K19/ISY/01) have completed partial fulfillment of dissertation titled "RETINAL VESSEL DETECTION USING RESIDUAL Y-NET" under my supervision in partial fulfillment of the MASTER OF TECHNOLOGY degree in Information Systems at DELHI TECHNOLOGICAL UNIVERSITY.

Guide's name: Prof. Kapil Sharma

ACKNOWLEDGEMENT

I am very thankful to **Prof. Kapil Sharma** (Professor, Department of Information Technology) and all the faculty members of the Department of Information Technology of DTU. They all provided us with immense support and guidance for the project.

My grateful thanks are extended to Delhi Technological University for providing us with the laboratories, infrastructure, testing facilities and environment which allowed us to work without any obstructions.

I would also like to appreciate the support provided to us by our lab assistants, seniors and our peer group who aided us with all the knowledge they had regarding various topics.

Place: Delhi
Date: June, 2021

Anindita Roy
Roll No. - 2K19/ISY/01
Department of Information Technology
Delhi Technological University

ABSTRACT

In this Corona Pandemic, Diabetic patients are affected a lot. Unfortunately, due to the consumption of steroids, people are affected by mucormycosis, which is a kind of fungal infection, making this situation worse. Patients get swollen red eyes, and diabetic patients are more vulnerable to it, as they suffer from Diabetic Retinopathy. It has become essential to determine the damage caused to the eye to save patients from vision loss. Only doctors can identify how the condition of the eye by physical examination. But, this is a tricky and time-consuming job. With the help of fundus photography and deep learning algorithms, the detection and classification process will speed up. There are many existing image detection algorithms, but they do not have efficient feature retention and lightweight architecture model. This paper proposes Residual Y-net architecture that works excellently on a balanced medium-size. With the help of segmented features it acquires reliable features which help in classification. It is a very lightweight architecture inspired by U-net, Deep Residual U-net, and Y-net. The addition of residual units in the network has significantly improved the accuracy rate. It is observed that a balanced dataset gives a much accurate performance than an unbalanced dataset. The proposed model's test accuracies on medium-size unbalanced and balanced datasets are 90.39% and 93.60%, respectively.

Keywords: CNN – Convolutional Neural Network, Res U-net – Residual U-net, Res Y-net – Residual Y-net

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

Diabetes has become an underlying condition in patients who are affected by the Coronavirus pandemic. Studies show that critical forms of COVID are seen in uncontrolled diabetic patients [1]. The process of healing in their body is slower than other patients that put them at higher risk. [2]Diabetic patients are mostly admitted to ICU (Intensive Care Unit) since they are most likely to have other health critical health conditions like hypertension, cardiovascular disease, etc. making them more vulnerable in this pandemic [3] [4] [5]. The most number of deaths are seen in the old age group with diabetes. However, various medications especially steroids made the situation worse and now people are struck with a fungal infection called, Mucormycosis [6]. Generally, patients are facing swollen red eyes. It is a dangerous infection taking many lives. It is again dangerous for diabetic patients as they suffer from Diabetic Retinopathy. A patient suffered from diabetic ketoacidosis [7] with right eye ptosis. Histopathology results disclosed the patient had mucormycosis. Even while treating for mucormycosis, the patient already [3] had diabetes. Fundus photography showed the presence of non-proliferative diabetic retinopathy. The medications used for Covid treatment are ocular toxic [8]. Few of them are responsible for inducing retinopathy in the patients. Therefore, the detection of diabetic retinopathy is still a critical situation for the medical industry.

Selection bias may emerge due to the limited sample size of this retrospective investigation. A large-scale research was required because this study was based on a single centre. Finally, the death rate for diabetic individuals with COVID-19 was 14.5 percent [9], which is significantly higher than the rate for non-diabetic patients. Diabetes, a high D-dimer level, and a low lymphocyte count at the time of admission were all risk factors for mortality in the hospital. As

a result, COVID-19 patients with diabetes require special attention.

COVID-19 has been linked to diabetic ketoacidosis [7] [10] in both T1D and T2D patients, as well as other severe illnesses. [11] Following the initial research from China that showed a high prevalence of diabetes ranging from 7.4% to 19.5 percent, numerous investigations found a greater incidence among individuals who required hospitalisation and/or died. According to Italian data, 35.5 percent of dead patients had diabetes, which is three times the rate in the general population. Diabetic patients required greater ICU care in Chinese cohorts, and the incidence of acute respiratory syndrome (ARDS) was 2.34 times higher. The death rate was significantly greater depending on the complications. A study in Wuhan of 191 patients showed an odds ratio of 2.85 compared to the non-diabetic population in diabetic patients.

While individuals with diabetes tend to have a higher chance of developing severe or critical forms of COVID-19, [12]the roles of diabetes, chronic hyperglycemia as measured by glycated haemoglobin (HbA1c), insulin insufficiency and/or resistance, obesity, and other comorbidities are yet unknown. Only one research examined the clinical manifestation of COVID-19 in individuals with and without diabetes (with or without comorbidities). [13]This Chinese investigation yielded some interesting results. For starters, persons with diabetes tend to experience milder symptoms at first. As a result, fevers were less common, perhaps delaying the first diagnosis. Second, a computed tomography scan of the breast indicated severe pneumonia infections in diabetic individuals. Third, diabetic patients (particularly those with no co-morbidity) have more evident biological problems, such as higher inflammatory biomarkers [such as C-reactive protein (CRP) and interleukin-6 (IL6)], elevated tissue enzymes [such as lactate dehydrogenase (LDH), and coagulation disorders (e.g. elevated D-dimer). These diseases, according to the authors, are linked to significant organ damage and a proclivity for thromboembolic events, as well as a "cytokine storm" characterised as a COVID-19 aggravating factor. Finally, lymphoma was more common and worse in diabetic individuals,

which is commonly cited as an indication of bad prognosis. Due to methodological limitations described below, these findings need to be verified by additional research, including data for other Caucasian and non-Caucasian groups.

COVID-19 mortality [5] varies from study to study, ranging from 2 to 15% in severe forms, over 20% and critical forms 50%. The overall death toll for COVID-19 in Belgium has been jointly reported to both patients and patients. At the time of writing, the mortality rate in Belgium has reached 15% of the selected individuals (68.7 / 100,000 population event); Forty-five percent of deaths occurred in hospitals and 55% in home care facilities. [9]Again, these numbers should be carefully interpreted, especially in home care facilities, where there are known limitations in screening and diagnostic tests as well as suspected but not yet confirmed deaths. By comparison, the death toll is 19% in France (37.2 / 100,000 cases), 63% in hospital deaths, and 37% in home care facilities. However, not all patients were diagnosed because of the low mortality rate of COVID-19. According to the European Center for Disease Control and Prevention, the current global mortality rate is 7 percent, and the overall diagnosis could further reduce these estimates.

Diabetes mellitus [13]and severe COVID-19 patients may be exacerbated or exposed to neuropathic symptoms and mild to moderate stress disorders. Sensitive analysis of the sensory neurons included in the defense tool provides a quick and easy way to truly measure and identify the type and distribution of these sensory defects. We recognize that the lack of history of diabetes mellitus and regular arterial development does not eliminate pre-existing diabetes mellitus. Long-term group study using cognitive measures, including QST and CCM, can help develop and improve developmental neuropathy associated with COVID 19 in patients with or without diabetes.

1.1. BACKGROUND

The eye is a spherical visual organ present in humans and animals that captures light from its surroundings and converts it into signals that the brain interprets as images. The brain can take information and alter abstractions from the outside world using these pictures. The eye is made up of several components, each of which has a distinct purpose. The retina, which is positioned on the inner rear of the eyeball, receives light from the environment after passing through the cornea, iris, and lens. The retina is made up of many layers, one of which includes light-sensitive rods and cones, and is connected to the optic nerve. Rods and cones that have been activated convey information to the brain.

Between the core, transparent vitreous body and the choroid is the retina. The vitreous body is filled with a transparent gel that changes composition concentrically, starting with a fluid phase in the centre and progressing to a fibrous network as it approaches the retina. The expansion of foot processing from the uterus to structural cells that support the retina's vascular and neuro-sensory properties creates this delimitation. The retinal pigment Epithelium, or RPE, is the outer retinal layer that extends from the centre of the outer eye. This is one of the blood barriers in the retina that has been linked to maintaining retinal homeostasis. Early DR lesions, or mild to severe instances, are typically located between the structural cells next to the spirit, which restrict the inner retina, and the outer retinal pigment epithelium (RPE).

The eyeball, like any other tissue in the body, has features that correlate to its division of work, such as size, colour, shape, and distribution. The retina of the eye has a distinct structural layout. The visible blood vessels are also the receptors. Central ocular occlusion and ocular occlusion are shown in the line emerging from the top right of the eyeball. The arteries of the

back and blood vessels are not really the arteries of the body. They are similar in shape to arterioles and venules in equal proportions with the walls of the vascular wall. The blood vessels that enter the eyeball are the branches of the optic nerves and nerves that rotate from the muscular and vascular system. The coronary arteries and veins are small branches of the optic nerves that migrate to the retina by penetrating the optic nerve, where they are in the middle until they reach the retina where they begin to branch.

These two blood vessels transport and release the retina, but the importance of retinal metabolism and catabolism also depends on the choroidal dilation of the uveal tract. The uveal tract is supplied by the branches of arteries, such as long or short ciliary blood vessels, and extends into the venous mucous veins, which drain the tissues of the dorsal part (like ciliary blood vessels). Small ciliated blood vessels pierce the uvea around the optic nerve of the eye. The uveal tract extends to the frontal lobe by the ciliary body, which also contributes to retinal homeostasis via blood vessels. If the RPE is yellow in the blood vessels of the uveal tract and the choroid is common.

Retinal function depends on balance and function of the uveal and peripheral nerves. There are exceptions to this, the most visible fovea or area is usually around the choroidal perimeter unless there are no vessels in this area. Fovea is in the segment for the optical disk and can be viewed from the center of the vertical line and below the horizontal line. These fovealies are in the center of the macula lutea, an area with xanthophyll pigment, this is also visible for other foveas.

1.1.1 DIABETES

Diabetes [14] is a disorder in the body because it is unable to produce or respond to the

hormone insulin, leading to abnormal metabolism and high blood glucose levels. Such levels of blood glucose can damage blood vessels and nerves, increasing the likelihood of developing other underlying diseases, such as diabetes mellitus retinopathy, in diabetic patients.
There are two kinds of diabetes [14]: type 1 and type 2. Type 1 diabetes is an autoimmune illness that causes the pancreas' insulin-producing beta cells to die, leaving the body unable to generate enough insulin to keep blood sugar levels under control. Because type 1 diabetes results in a decrease of insulin production, insulin must be given on a regular basis. Type 2 diabetes is a metabolic condition that causes hyperglycemia, or high blood sugar levels, as a result of the body's inability to efficiently use or generate enough insulin. Type 2 diabetes is defined by the body's inability to metabolize glucose, resulting in excessive blood sugar levels that damage bodily organs over time.

Diabetes is a disease that affects everyone, and the odds of getting diagnosed with it grow with age. Attempts to determine the extent (prevalence) of diabetes in the population range from 1% to 6%. Half of these patients are undiagnosed and hence unable to access health treatment.

Prevalence of diabetes

According to the International Diabetes Federation (IDF), [16]463 million people worldwide would have diabetes by 2020, with 88 million in Southeast Asia. 77 million of the 88 million population are Indian. The prevalence of diabetes in the population is 8.9%, according to the ID card. After the United States, India has the second largest number of children with type 1 diabetes, according to the IFF. It also has a significant impact in the development of type 1 diabetes in coastal youngsters. Diabetes is responsible for 2% of all fatalities in India, according to the World Health Organization.

[16]In India, the number of diabetics has grown from 26 million in 1990 to 65 million in 2016.

According to the Ministry of Health and Family Welfare's 2019 National Diabetes and Diabetic Retinopathy Survey Report, the illness affects 11.8 percent of persons over the age of 50. The DHS study found that the prevalence of diabetes in adults under 50 was 6.5% and that of pre-diabetes 5.7%. This increase was similar for men (12%) and women (11.7%). It was mostly in the suburbs. A study of diabetic retinopathy found that 16.9% of people over the age of 50 were affected by diabetes. According to the report, diabetic retinopathy is 18.6% in those aged 60-69, 18.3% in those aged 70–79, and 18.4% in those over 80 years. In the 50–59 age range, the frequency was 14.3% lower. Diabetes is expanding in commercially and economically developed countries such as Tamil Nadu and Kerala, where numerous research institutes are also growing.

Diabetes was diagnosed in 6.6 percent of North American inhabitants (Harris et al., 1987), and half of them had a pre-existing disease. The remaining half had diabetes, with roughly a quarter of patients requiring insulin. Insulin usage is linked to type 1 diabetes, and it's no different for type 2 diabetes. Diabetes affects people differently as they become older. The frequency is less than 2% between the ages of 20 and 44. Diabetics in the 55- to 74-year-old age group make up around 15% of the population, and the prevalence is projected to rise as people become older due to increased morbidity, demographic shifts, and decreased diabetes.

1.1.2 DIABETIC RETINOPATHY

Diabetic Retinopathy [17] is caused by blood-retinal disorders. The molecular biology of retinal physiology has yet to explain these variables. Alterations in surface adhesion, structure, and elasticity occur as a result of these chemical changes spreading to the retina and capillary cells. High glucose levels (hyperglycemia) and the damaging effects of stress products on metabolic pathways are the causes of this disorder. Structure, motor, and mediating activities are all

present in damaged proteins. As a result, people with early diabetes may have a variety of physical symptoms of eye illness. The lack of lymphatic or extracellular fluid systems, which tend to compensate for the retina and contribute to more severe symptoms, is thought to worsen these consequences.

Type I [18] diabetes is caused by autoimmune or viral loss of pancreatic B cells, which leads to insulin insufficiency. Type 2 diabetes has less well-defined causes. It's thought to have a negative effect and produce insulin. These two forms of diabetes have comparable retinal consequences. The differences in the consequences are due to the fact that the patients were diagnosed with diabetes at various dates. Type II diabetics frequently have subclinical illness before seeking medical help, while type I diabetics' symptoms might be seen in other organs.

The increased viscosity of diabetic blood, which exacerbates hyperglycemic damage to endothelial cells and the basement membrane, is one of the many repercussions of hyperglycemia. The permeability of the vessel walls is affected by the presence of fenestra and the lack of thin connections. Because endothelial and pericellular cells are capillary cellular or comprised of basement membrane solitude, this toxic or biochemical assault might cause them to die completely.

The capillaries are just a partial structural component of pericytic narrowing cells, which are thought to be regulatory. Their loss, along with the loss of smooth muscle mass in the retina's bigger arteries, can raise hemodynamic pressure, resulting in capillary distention and accumulation, also known as micro-aneurysms. [19] Damaged endothelium, motile and inflammatory blood cells are all involved, and it clings to them. This causes capillary occlusion, and MAs may indicate an increase in new vascular abortion as a result of attempting to re-channel or enhance the blocked vessel. MA is one of the earliest visible alterations in micro-vascular morphology. The beginning steps for DR development include pericardial loss,

endothelial cell injury, and MA formation (possibly linked with decreased RPE function).
Loss of vision or total blindness can result from changes in the function and structure of the micro- and renal blood vessels. The progression, prognosis, and symptoms of a ruptured retinal blood barrier varies from person to person. Some or all of the usual lesions of early retinopathy may not occur in a healthy diabetic retina. [17] There may be lesions, such as MA, that appear to be important in the progression of DR but do not present or are undetected in the early stages of the disease. Despite these inconsistencies, the language used to describe the condition and phases of DR is linear. It begins with a healthy or normal retina and progresses to a stage of extensive scarring and retinal tissue loss. Between these two extremes, the development of DR typically falls into one of two groups. Non-proliferative and proliferative DR, or NPDR and PDR, are the two types of DR.

It causes damage to the retina's tiny blood vessels. Retinal blood vessels can burst, rupture, or become clogged as a result of the side effects of diabetes, impairing the delivery of nutrients and oxygen to portions of the retina, resulting in visual loss over time. Such structural changes may initially lead to blurred vision and in the final stages, even retinal detachment and / or glaucoma. It results in swelling of blood vessels in the retina of the eyes. It gives mild issues in the vision. But, if not cured early, the patient may lose his or her eyesight [9]. This situation of the eyes of the patient is coined as Diabetic Retinopathy. Harm to veins in the retina happens when the veins get hindered because of high sugar levels in the blood. New blood vessels attempt to develop but are not able to develop appropriately. These create small lumps in the blood vessels, which may now and then hole blood into the retina [10]. The width of the veins changes and the state of the retina gets highly severe. Ophthalmologists have specified different forms of damages caused to the retina of the eyes.

i) Microaneurysms [11]:- These are the red spots that happen in the eye chunk of diabetic patients. Even though they don't influence eye vision, they help in distinguishing the beginning

period of diabetes.

ii) Cotton-fleece spots [11]:- They are white and feathery fixes in the eye. These appear when the eye nerve strands damage. There are associated impacts, too like, haemorrhages and microvascular infarcts.

iii) Hemorrhages [11]:- This shows up as splendid red or dim red fixed in the eye. They can seep inside in the eye. The dividers of the veins may break in the retina.

iv) Exudates [11] :- These happen on the outer side of the retina. They show up as sparkly yellowish in shading. They have sharp edges. Indeed, even microaneurysms, which have no bloodline in them, are additionally called hard exudates.

Different stages of diabetic retinopathy:

Mild non-proliferative diabetic retinopathy

[18]This is the first stage of diabetic retinopathy, characterized by small areas of inflammation of the blood vessels in the retina. These inflamed areas are known as microaneurysms. Small amounts of fluid can penetrate the retina at a stage that triggers inflammation of the macula. This is an area near the center of the retina.

Moderate non-proliferative diabetic retinopathy

[21]More lesions appear at this stage because more capillaries feed the damage to the retinal tissue and the retina becomes more ischemic (lack of blood flow, hence hypoxia)

Severe non-proliferative diabetic retinopathy

[18]At this level of DR, many blood vessels are affected. Blood vessel supply of oxygen to the retina is severely compromised due to accumulated vessel damage. When this occurs,

certain areas of the retina start sending biochemical signals to the body that they need oxygen. Increased swelling of small blood vessels begins to interfere with blood flow to the retina, preventing proper nutrition. This causes blood and other fluids to build up in the macula.

Proliferative diabetic retinopathy

[18]This is the advanced stage of the disease, in which new blood vessels form in the retina. Since these blood vessels are generally very fragile, the risk of fluid leakage is higher. This can cause various vision problems, such as blur, reduced vision, and even blindness. In response to the need for oxygen, new vessels (neovascularization) begin to grow within the retina. These new vessels are an aborted attempt of the retina to regain its oxygenation need, but these vessels are compromised and fragile. These vessels "break" easily causing severe bleeding into the vitreous gel of the eye and therefore, causing consequent loss of vision. Also, these new vessels can attach themselves into the vitreous gel and cause traction on the retinal plane, causing retinal detachments.

FIGURE 1: NO DIABETIC RETINOPATHY EYE

FIGURE 2: MILD DIABETIC RETINOPATHY EYE

FIGURE 3: MODERATE DIABETIC RETINOPATHY EYE

FIGURE 4: PROLIFERATE DIABETIC RETINOPATHY EYE

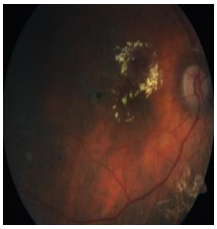


FIGURE 5: SEVERE DIABETIC RETINOPATHY EYE

Assessment of retinal health

1) *Ophthalmoscope*

[19]It damages the small blood vessels in the retina. As a result of the side effects of diabetes, retinal blood vessels might burst, rupture, or become blocked, limiting the supply of nutrients and oxygen to parts of the retina and resulting in vision loss over time.

A visual field, often known as a perimeter test, assesses the examinee's ability to see straight and peripherally. The goal of this test is to see whether there are any regions around the cigarette that create blind spots.

2) *Fluorescein angiography*

[19]The doctor can check the retinal blood vessels using fluorescein angiography. Injection of a vegetable-based dye into the patient's bloodstream. A sequence of fast, successive pictures of the eye are captured as blood flows through the retina. These photographs provide important details regarding his condition. One of the most significant tests for determining the diagnosis and therapy of retinal diseases is fluorescent angiography.

A high-powered camera with interference and fence issues is required by the FA. To generate a contrast image of the early phases of the angiography, fluorescein is administered into a transfusion, generally via a blood vessel standing at a high pace. The blue fashion has passed through white light from the light source. The fluorescein-free molecule absorbs the blue light (flash 465–490 nm), and the microscope emits light with the longest wavelength in the fluorescent, yellow-green spectrum. The wavelength ranges from 520 to 530 nanometers. Only light produced by high fluorescein is caught by the 520-530nm barrier filter. After the injection,

images were collected for up to ten minutes, depending on the pathology that was triggered. Digital or 35mm film was used to capture the images.

The pigment travels via the short posterior veins and arrives on the optic nerve and choroid within 8-12 seconds after injection into the uterine vein. This is dependent on the patient's age and cardiovascular system's condition, as well as the degree of colour injection. When the choroid lobes fill, the choroidal circle fills, causing a choroidal flush, a unique, dramatic hyperfluorescence. After 1-3 seconds, the retina rotates (11-18 seconds after injection). The filling of the retina's arteries, arterioles, and capillaries is referred to as the early arteriovenous phase. As the colour fills the veins in the laminar structure, a late arteriovenous or laminar venous phase follows. The non-capillary region of a typical macula appears black because to xanthophyll pigment blocking choroidal fluorescence and numerous epithelial cells packed with retina. The peak phase and maximum brightness last around 30 seconds, followed by repeated phases. Approximately 3-5 minutes after fluorescein injection, the rehabilitation phase begins. Fluorescein does not appear regularly on retinal arteries after 10 minutes, but it spots and penetrates numerous tissues, including the optic nerve, the Bruch membrane, and the sclera.

Changes in the FA [22] are sometimes referred to as average changes. Hypofluorescence refers to a reduction in predicted fluorescence, whereas hyperfluorescence refers to an increase in expected fluorescence. [7] Hypofluorescence can develop as a consequence of vascular injury or as a result of interference. Earaches, hoarseness, diabetes, and atherosclerosis are just a few examples. Retinal haemorrhage, subretinal fluid, or a variety of abnormalities in the retinal epithelium, such as those seen with lipofuscin in illnesses like Stargardt's disease, can all inhibit choroidal fluoride. The tissue's typical fluorescence may be absent or delayed as a result of total vascular injury. Occlusion of the retinal or choroidal arteries, as well as the posterior silary artery that connects to the optic nerve, might cause this. Hyperfluorescence can be caused by

fluorescein leakage, staining, pooling, infection, and treatment such choroidal neovascularization or retinovascularization, as well as recurrence. colourful epithelium that does not prevent the choroid from releasing fluorescein. The borders of areas that do not fit into the FA display are growing and bamboozling. This is incompatible with the facility's design. More flaking occurred throughout the angiography as a result of the abrasion, although the margins were just altered. Fluorescein can stain normal patterns like optic nerve ends and sclera, but it may also stain diseases like drusen and dysiform scars. When fluorescein progressively fills the liquid-filled region, it forms a pool. When a layer that typically flashes fluorescence vanishes, this is known as transmission or window damage. This generally happens when the RPE is absent, resulting in early FA viscosity. The applications of lost fluorescence and edges are still varied. When active fluoresce models like optic nerve drusen and lipofuscin are used, autofluorescence can be detected before fluorescein ink injection. Lipofuscin fluorescens can be used with some specialised equipment like as laser scanners and fundus cameras. Some special equipment equipped with laser scanners and fundus cameras can use lipofuscin fluorescence to record the health of the RPE layer.

FA can lead to a variety of problems. A brief rash, which occurs in 3–15 percent of patients, vomiting (7 percent), and itching are the most frequent reactions. Urticaria, pyrexia, thrombophlebitis, and syncope are more severe responses. On staining, local muscle necrosis can occur, but moderate and red discomfort is frequent. Anaphylaxis, heart disease, and bronchospasm are all serious life-threatening events that might occur, although they are extremely infrequent. The death rate is expected to be 1 in 221 781. [3] Although no adverse effects have been documented during pregnancy, this is a contraindication. [4] [5] Although malignant cases have the lowest occurrence, prior consent should be acquired before fluorescein angiography. In addition, imaging equipment should be well-equipped and ready to handle the complications associated with fluorescein angiography.

3) *B-Scan Ultrasound*

B-scan ultrasonography [20] examines the back of the eye using high-frequency sound waves. This technique offers a comprehensive image of the eye that may also be used to assess the retina's state. A cross-sectional picture of the interior tissues of the eye is typically obtained using ultrasound and a B-scan. It is often used to measure tumors or to detect retinal or choroidal secretions. It is also used when the focus on the retina is cloudy and not clearly visible, for example in patients with conjunctival hemorrhage, conjunctivitis or trauma. The technology is very fast to handle, non-invasive and completely safe.

Ocular ultrasonography, commonly known as ocular echography, "echo," or "B-scan," is a quick, non-invasive diagnostic used in clinical practise to evaluate the anatomy and health of the eye. Optimal noninvasive information can be provided by accurately visualizing facial tissue, and is particularly useful in patients with obstructed or obstructed or obstructed epilepsy by ophthalmoscopy (e.g., ophthalmoscopy). Large cornea, ductal artery, or hemorrhage).

Some researchers also use a highly skilled visual ultrasonographer to perform visual ultrasounds during business hours around the clock. Therefore, surface researchers may not have technicians experienced in visual ultrasound. This weakness can be seen in patients who are seen after hours, while on the phone. The skill of using a square ultrasound is an invaluable asset for dentists who want to quickly, remember, and not see the world and test patients accurately. Remember, where an open world injury occurs, the echographer should arrange it, as eye pressure will always be fatal. Here, we present a simple “life care” guide for eye surgeons using ultrasound.

Ocular Ultrasound can evaluate the entire world in only five manoeuvres: four dynamic quadrants and one static portion on the macula and optical disc, commonly known as the longitudinal macula (LMAC). T12, T3, T6, and T9 are the appearances of quadrants. One hour

of the eye corresponds to these numbered rectangles. T12 is the superior quadrant of the eye, T3 is the right eye's nasal quadrant (temporary quadrant of the left eye), and so on.

4) *Fundus Photography*

The fundus photography [20]uses special cameras to document and detect the progression of certain retinal diseases, such as diabetic retinopathy, as well as to monitor its treatment. Fundus Photography is a technique for photographing the inner area of the eye. This technique allows visualization of the main structures shown in the posterior interior, such as the central and peripheral retina, theoptic capillaries, and focal points. Such images are identified as the central lateral retina, the macula (the darkest part in the center), and the optical disc (the white spherical structure in the retina). With this type of photograph, it is possible to identify places and species and, if possible, draw results from them.

The fundus camera's visual design is based on the one-eyed indirect ophthalmoscopy principle. [20]The fundal camera shows the fundus in an upright, enlarged position. The basic camera views 30 to 50 degrees of retinal area at 2.5x magnification, and enables some adjustment of this ratio via 15° zoom or an extra lens that offers 5x magnification, up to 140 degrees with the wide-angle lens, which reduces the picture in half. Indirect ophthalmoscope optics are comparable to fundal camera optics in that the viewing and illumination systems use separate pathways.

Observing light travels via a series of lenses through a donut-shaped hole, then through a centre hole to form a ring, before travelling through the camera lens and through the cornea to the retina. The lighting arrangement creates a donut in which light reflected from the retina travels through lit holes. Because the two systems' light pathways are separate, there is very little reflection from the light source in the prepared image. At low strength, the image-forming beam continues towards the telescopic eyepiece. When the camera button is hit, the mirror

obstructs the lighting system by allowing flash light to travel through the eye. A mirror falls in front of the observation telescope at the same time, directing the light onto the reception media, whether film or digital CCD. The exit seals must be positioned to create a focussed picture on the receiving medium due to the eye's inclination to accommodate when gazing through the telescope.

1.2. Understanding The Problem

Medical Science uses high resolution systems to capture the condition of the eye. Doctors have to examine the length, width, branching of blood vessels to carefully examine the disease and to how much extend it has spread. But, this is a tedious and time-consuming job. In future, there will be many variations in the diseases as well. Therefore, it becomes necessary to create automated approach which will assist practitioners to detect diabetic retinopathy through pictures taken.

Blood vessel extraction is an important step, since a slightest difference in the structure of vessels may result into different diagnosis. Furthermore, the possibility of bizarre cases demands for a superior level of feature extraction. Deep Learning techniques are preferred to use as it has varied layers to extract features from the images. Convolutional Neural Network is one of the deep learning algorithms, popularly used for image processing. CNN requires less pre-processing because of which one can easily pass the raw input images through it. It reduces the input images in such a way that all the important features are retained particular for predictions.

However, CNN [23]wants large number of training data that also labeled. It takes a lot of time for training the model. Transfer Learning [24] is an excellent alternative, when there are less

publicly available labeled data. There are very few datasets which are publicly available. Transfer Learning helps to train model using a already trained model on a bigger dataset.

But, with the limited available resources and dataset, it becomes important to find an optimal solution which works on small or medium size dataset along with limited resources. The model might under-fit when a model runs on a small dataset. [25] While in large datasets, the unbalanced class distribution causes the model to overfit. The model will undergo training securely if the dataset is present in similar class distribution, i.e., the number of images in each label is almost the same. For this issue, there is a need of such a machine learning architecture which works fine with small or medium dataset, while does classification and detection desirably.

There is also issues of loss of information when goes from one layer to another in the neural architecture. Feature loss will lead to wrong detection and classification. Requirement of such a deep learning model which will not only work on small or medium datasets but also retains information about the images is the problem statement.

U-net [12] is a type of CNN specifically invented for biomedical image classification. It works fine when the dataset is small. The architecture helps to train end to end, and it classifies each pixel in the image into a class label. It is better than patch-wise classification because it is redundant and results in loss of information. In contrast, pixels help in better label classification because it is not redundant.

U-net, Deep Res U-net [13], and Y-net [14] inspired the proposed model. The VGG network serves as the backbone for U-net and Y-net, whereas Res-Net is the backbone of Deep Res U-net. U-net has convolutional blocks that contain a stack of convolution and pooling layers that are responsible for the degradation of feature details. To rectify this problem, Deep Res U-net

was proposed which used Residual units [15]. It uplifts the accuracy of neural network. The residual unit is a combination of the input layer, convolutional layers, batch normalization, and activation functions. From each passing layer in a residual unit, the neurons tend to lose a few of the features of the actual layer, so when the original input layer is added with the output layer, then all the lost information along with the processed output can be obtained. To solve the same problem of feature loss, instead of using conventional encoder-decoder architecture, Y-net applies two encoders and one decoder. It adds the output of the convolutional block of the first encoder with the output of the corresponding convolutional block of the second encoder. The proposed model is termed as Deep Res Y-net because it uses residual units [15], and has Y-net architecture. The primary purpose of this method is to contain as much information as possible. It reduces the possibility of degradation of the model, which generally happens when going from higher to lower levels. From each block of the encoders, features are combined and moved to the appropriate stage of the decoder. This help in filling the gaps when any information from the neurons is corrupted or lost at any point.

This thesis has five sections. The first section is an introduction to the work, which explains the context of the project, the proposed methodology and dataset information. The second section is a glimpse of work that has been done by various authors in the past years on Diabetic Retinopathy using deep learning architectures, about the dataset they used and results they obtained. The third second goes over U-net, Deep Residual U-net, Y-net, and the proposed Deep Residual Y-net model in great detail. The fourth section explains the dataset, experimental setup to execute the model, and results in graphical forms observed after execution. The final section concludes this paper on its proposed method and future work.

CHAPTER 2
RELATED WORKS

There has been a considerable amount of work in diabetic retinopathy classification. Many deep learning methods are applied. This paper [16] uses DenseNet over 3050 images obtained from Kaggle. It gives a validation accuracy of 84.10%. It describes how a skewed dataset overfits the model and uses the kappa score to measure the model performance due to the hardware constraints and skewness of the dataset. Integrated Shallow Convolutional Neural Network, on datasets ranging from 200 to 35000 photos in quantity, was used [17]. An integrated system helps to bring out the best part from different neural architectures. This [18] paper uses an ensemble learning approach. It boasts strong base learners to put a competent system forward. It executes 1200 group experiments, in which ensemble learning excels. Influenced by this, many authors integrated various deep learning models to perform image detection. In the following proposed architecture [19], a combination of three convolutional neural architectures (SE- ResNeXt50, EfficientNet-B4, and EfficientNet-B5) is used, then applied to that transfer learning. APTOS2019 Kaggle dataset is used in this paper. It contains 3662 training images in similar class distribution. 0.8265 of accuracy obtained from this method. MobileNet V2 is a deep CNN that has been designed especially for computer vision and smartphone applications. It is a lightweight network. [20] Upon this, fine-tuning is applied, due to which accuracy reaches 91%. The dataset comes from Kaggle, which has 3662 images in total.

U-net is also used for image classifications. For nuclei detection, this paper [21] compared two commonly used CNN methods (CNN1 and CNN2) and Residual inception channel attention(RIC) U-net on Cancer Genomic Atlas dataset. Compared with original U-net and RIC-U-net, they got f1 score of 0.8155 and 0.8278. The In the paper [22], weighted residual U-

net has been used in which skip connections are utilized. The method adds the input weights into the network at every level to extract features. The experiment is performed over DRIVE and STARE datasets. The accuracy seized in this process is 0.9655 on the DRIVE dataset, and 0.9693 on the STARE dataset. On the same dataset, [23] the authors in this paper proposed a bridge-style U-net. The approach is to add the nearest feature output from one block into the next block as its input information. [24] [25]The closest features rapidly help in gaining the most relevant information. DRIVE, STARE, and TONGREN datasets were used to train and test the proposed model, where 0.9567, 0.9658, and 0.9652 of accuracy had been recorded over the three datasets, respectively. The spatial attention module is designed for detection and classification, is a part of the convolutional block attention module [26]. It is used [27] to improve the performance of the U-net. This module calculates a spatial attention map using the spatial features. Then, it applies global and average pooling layers over the channel axis to produce a defining quality. It is tested over DRIVE and CHASE DB1 dataset in 100 epochs reaching an accuracy of 0.9618 and 0.9905, respectively. [28] This paper used the SLIC superpixel algorithm, which is a patch generating algorithm, to solve the problem of an imbalanced dataset. In U-net, the authors applied residual connections and inception modules on the IDRiD dataset and obtained 97.95% accuracy. [29]CAR-Unet is introduced in which a modified channel attention network is developed. An updated channel focus network replaces the original blocks. This approach employs layers of max pooling and average pooling. These modifications in U-net helped to get an accuracy of 0.9699, 0.9751, and 0.9743 on DRIVE, STARE, and CHASE datasets, respectively. There are combinations of transfer learning architectures with U-net. In this paper [30], the authors replaced the usual convolutional blocks of the U-net with the VGG16 network in the encoder part. Instead of using completely connected layers, the Global Max-pooling layer is put after the 5th convolutional layer. Instead of a sigmoid activation function, the last layer in the decoders uses a softmax activation function.

The majority of the research related to the U-net and its modifications are applied on the

DRIVE, STARE, and CHASE datasets, which are tiny dataset, as can be seen from the previous found works. The experiment in this paper tries to research upon a much bigger dataset.
Y-net is also recently used in various fields of medical imaging. Y-net [14] has been applied over breast biopsy images for segmentation and classification. It outperforms the traditional encoder-decoder structure by 6%. It based upon the VGG11 encoder. [31]It is added with augmentation to 1000 chest x-ray images. The results were based upon the acceptable percentage that came out to be 95.8%. Then, this architecture has experimented upon polyp detection [32]. It gets compared with U-net and pre-trained U-net. It gives an 84.4% of recall and 85.9% of f1 score, which is way better than the other two models. Although, the precision it obtained is 87.4, which is less than U-net and pre-trained U-net. There is someplace for improvement in the architecture because it performs well on few classification parameters.

This paper used connected convolutional AlexNet [46]. Upon four database has been tested, which are DRIVE, STARE, CHASE DB1 and HRF database. The images are chopped into 50X50 for increasing the quantity of the images and for easier blood vessels extraction. The new model is obtained from the pre-trained AlexNet network. 30 training epochs with learning rate set as 0.0001 is applied. Accuracy, specificity and sensitivity is observed for all the four databases. On single database set, the method obtained an accuracy of 0.9968 and on cross dataset, it obtained 0.9591. InceptionV3 network [47]has been used, which gives an accuracy of 48.2%. It used Kaggle dataset produced by EyePacs. It shows improvement in accuracy when the number of images are increased with rise of pixel from 200 to 500. InceptionV3 is also used without data augmentation on kaggle dataset where 48.2% of accuracy is attained. It does binary classification of images. [48]The dataset is gotten from Eyepacs clinic in United States and three emergency clinics in India. The calculation utilized is profound learning. For each picture evaluation of seriousness is determined and contrasted and the preparation set evaluation. At that point, the rule of computing seriousness esteem is changed in accordance with decline the blunder. This procedure is imitated commonly on a similar picture

so the calculation can take in the seriousness esteem from the power of the picture pixels. It is done over all the pictures present in the preparation set. This paper utilizes convolutional neural system. The restriction to this framework is the neural system is just prepared with explicit evaluating without giving any genuine definitions to those reviewing, so it can't distinguish whether the picture portrays microaneurysms, hemorrhages, etc.

Apart from the development in the classification part, there has been research over the cleasing of image dataset as well. Like in [49] thresholding is a procedure of changing over dark scale picture into parallel picture. This strategy fixes a limit of pixel esteem. Whichever pixel surpasses that limit is given worth 1. Along these lines the highlights are removed. At that point at long last morphological cleaning is finished by expelling salt and pepper clamor utilizing middle channel. It gives great outcomes because of the twofold picture improvement. The impediment is that in certain pictures it can't recognize the availability, in light of this it is giving wrong division. There are other researches for image enhancement. [50]Performs morphological tasks to remove highlights of the optical plate in the natural eye image. It performs disintegration and afterward enlargement. This assists with expelling commotion from the fundus picture. It expels gaps from the picture. At that point the picture is changed over into dark picture. This assists with recognizing objects in the picture all the more unmistakably. Presently, to distinguish round items like blood clusters, and so on are recognized utilizing Circular Hough Transform. At last, to section the influenced territory, dynamic shape model is utilized. Shapes are brought into the picture. The district based dynamic form is generally reasonable for discovery of items and its highlights in the picture. The pictures procured can be boisterous, have no consistency, there can be distinction in the brilliance and contrast in the complexity. This will prompt error in preparing the robotized calculations. The availability in the models will be lost. In this way to unravel this, picture upgrade is finished. Picture Enhancement incorporates:- Sharpening the pictures, De-obscuring out of center pictures , Featuring the edges, Improving picture contrast, Lighting up image, Expelling commotion ,

Change into dim scale picture, whenever required. All these are acquired by applying different sorts of channels like mean channel [51], median channel, Gaussian channel, butterworth channel, and so on. Histogram balance [52] is likewise performed for differentiate improvement. For splendor adjustment, a steady worth is included or deducted from all the pixels in the picture. For differentiate alteration, a steady worth is duplicated by all the pixels in the picture. This is additionally one sort of scaling of picture. For edge sharpening, spatial channels are applied. There are Laplacian administrators, which are second request subsidiary, high lift channels, un-sharp concealing, slope, and so on for this. [53]High recurrence pictures and low recurrence pictures are gotten. They are deducted from unique picture as indicated by the need, to get sharp edges. For evacuating clamor, the mean filter is viewed as best. It neatly evacuates the salt and pepper commotion, which happens because of unnecessary improvement of the picture. Clamors can be smoothed edges or obscured picture districts. In this way, clamor models are applied over it to evacuate commotion.

CHAPTER 3
METHODOLOGY

3.1 DEEP LEARNING

Machine learning [54] technology is used in many parts of modern life, including web search, internal social network filtering, and offers on 'e-commerce' websites, as well as consumer items like cameras and phones. The machine learning system is used to recognise item pictures, convert speech to text, adapt to new objects, articles, or goods, as well as the requests of users, and pick appropriate search results. Deep learning is a type of technology that is increasingly being used in these systems. Deep Learning is a branch of artificial intelligence. Profound learning is altogether worried about calculations enlivened by the structure and capacity of fake neural systems which are motivated by the human mind. Deep learning is utilized hardly lifting a finger to anticipate the unpredictable. Deep learning, additionally called a subset of AI which is an expert with a very intricate range of abilities so as to accomplish far superior outcomes from similar informational collection. It simply based on NI (Natural Intelligence) mechanics of the organic neuron framework. It has a perplexing range of abilities in light of techniques it utilizes for preparing for example learning in profound learning depends on "learning information portrayals" instead of "task-explicit calculations." which is the situation for different techniques.

Traditional machine learning techniques have limited capabilities in processing natural data in its raw form. Engineers and field skills have been necessary for decades to create feature extractors that transform input (such as pixel values in pictures) into internal representations in pattern building systems or machine learning systems. The learning subsystem (typically the

classifier) is able to recognise and categorise patterns in the input. The technique of representation training allows a machine to examine raw data and locate the representatives required for detection or distribution. A model of representation with several layers of representation is an in-depth research. It's made up of basic but unusual combinations. Each model transforms as a representation of one level (starting from the first instruction) to a higher and higher level of understanding. By adequate connections such as transitions, very difficult tasks can be realized. For the division of labor, higher-level representatives focus on strategies that are critical to diversity and prevent inaccurate exchanges. An picture, for example, is a pattern with pixel values, and the qualities learnt in the first representation are frequently substituted by or without particular boundaries and areas in the image. Regardless of any little modifications in the working edges, the second layer generally discovers a motor by visualising the arrangement of the edges. The third layer can transform motifs into bigger textures that fit the normal material's shell, while the outer layer captures the material as these parts are linked. The main idea of the in-depth study is that the layers of the structure were not created by human engineers: they were learned from the data through a general learning method.

Deep learning is a collection of basic models that are all (or almost all) learnt and that frequently calculate non-linear input-output grids. Each model in the group alters the format, enhancing the representation's variety and ambiguity. Using many non-layered layers (e.g., depths of 5 to 20) allows the system to access information that is not sensitive to mass and does not vary, such as backdrop, motion, illumination, and surroundings, while also utilising the material's extremely complicated properties.

The backpropagation method for computing the slope of an objective function with respect to the stack weight of a multilayer module is just a practical implementation of the chain rule for derivatives. Working backwards from the slope with regard to the output of a module, a target's derivative (or slope) may be computed with respect to the input of that module (or subsequent

module input). By recursively applying the back-division equation to all modules, we may multiply the gradient. The gradient from the top (where the network makes predictions) to the bottom can be multiplied (where the external input is fed). It's simple to compute the gradients in proportion to the weight of each module after these gradients have been established.

Reverse neural network architectures are used in many deep learning applications to learn how to translate a fixed-size input (such as an image) to a fixed-size output (e.g., probability for each of several categories). A set of units calculates the weighted sum of its inputs from the previous layer and sends the result to a nonlinear function to move from one layer to the next. The rectified linear unit (ReLU), which is just a half-wave rectifier f (z) = max, is now the most used nonlinear function (z, 0). Smooth nonlinearities like tanh (z) or 1 / (1 + exp (-z) have been employed in neural networks in recent decades, but ReLU learns considerably quicker on multilayer networks, allowing the construction of more complex models. Hidden units are units that are not at the input or output level. Hidden layers can be thought of as nonlinear input distortions that allow the categories to be linearly separated from the final layer.

Back propagation and neural networks were mostly disregarded by the machine learning community. It was likewise disregarded by the computer communities' visual and speech-recognition communities. Learning meaningful multi-step extracts with minimal prior information was considered to be impossible. A basic slope computation, in example, was considered to be locked in weak local minima - heavy settings where even minor modifications would not lower the mean error. A bad local minimum is usually an issue in many networks in practise. Regardless of the starting conditions, the system usually always produces high-quality solutions. The local minimum is not a major concern, according to recent theoretical and theoretical conclusions. The scene, on the other hand, is encased in a tangle of numerous saddle points, each with a gradient of zero and a top curve in most dimensions and curves in the rest. The study appears to suggest that the saddle points are many, with just a few curve orientations

at the bottom, but nearly all of them have the same amount of objective movement. As a result, it doesn't matter where of these saddle points the algorithm is gathered in.

Then there was a surge in interest in deep forward learning. Uncontrolled learning techniques were developed by the researchers, allowing them to build layers on detectors without requiring tagged data. In order to simulate the performance of function detectors (or raw inputs) in the layer below, objective learning of each layer of function detectors has to be able to reconstruct. By pre-training the detector layers of multiple increasingly more complicated functions, the original network's bulk may be reduced to a manageable level. After that, the last layer of output units could be added to the network's top layer, and the complete deep system could be set using standard backpropagation. This was notably useful for detecting handwritten numerals or identifying pedestrians when the quantity of data indicated was restricted.

Speech recognition was the first significant use of this pre-training approach, which was aided by the development of powerful graphics processing units (GPUs) that were simple to design and allowed researchers to train networks 10 or 20 times quicker. This approach was used in 2009 to convert a collection of probabilities of distinct speech fragments that may be represented by a frame in the centre of the window from a short time window of coefficients derived from sound waves. It set new records in conventional voice recognition benchmarking tests with a short vocabulary, and it was soon improved to hit new highs in huge vocabulary tasks. Since 2009, several large voice teams have been using the Deep Web version, which has been implemented on Android phones. When the number of tagged instances is limited or when there are numerous examples in the transfer settings for some "source" activities, unsupervised pre-training helps to minimise over-fitting, but few Generalization benefits are considerably better when applied for specific "target" tasks. It turns out that once deep learning is stabilised, only small data sets require a pre-training phase.

However, unlike networks with full connection between adjacent layers, there existed a form of deep advanced network that was considerably easier to train and generalised far better. The convolutional neural network was responsible for this (ConvNet). It has had a number of practical triumphs at a period when neural networks were underutilised, and it has recently gained widespread acceptance in the computer vision field.

3.2 CONVOLUTIONAL NEURAL NETWORK (CNN)

CNN [55] is a collection of networks just like neurons in Human brains. Their main objective is to analyze the patterns in an image using the multiple neurons with weights at every CNN. CNN puts three methods together which are scale, shift and distortion. Covnets are used over feed forward network as they perform better in fitting images in the neural network by reducing the spatial size and number of learning parameters. The very first step of convolution layer is the kernel. It is a filter matrix which runs over the input image. This means they are multiplied with each other resulting into a feature map. Kernel helps in sharpening, detection, blurring, etc. providing a filtered input image. Every node in a layer is directed by an activation function. This function governs the output at every layer. Linear, tanh, rectified linear unit (ReLU), sigmoid are popular activation functions. Then, there are pooling layers. These are used for non-linear sampling. Non-linear sampling is faster in converging characteristic features.

ConvNets are built to handle data in many fields, such as a colour picture made up of three 2D fields representing the strength of pixels in three colour channels. There are various types of data modalities: 1D for signals and outcomes, such as voice; 2D for pictures or sound spectrograms; and 3D for video or volume images. ConvNets are based on four key concepts that take use of natural signal properties: local connection, weight loading, aggregation, and the

utilisation of many layers.

A typical ConvNet [39] architecture is developed in a number of stages. There are two types of layers in the initial stage: convolutional layers and meeting layers. The convolutional layer's units are arranged into function maps, with each unit linked to the function maps of the preceding layer in local patches by a collection of weights known as a filter bank. Nonlinearity, such as a ReLU, is used to convey the outcome of this local weighted sum. The filter bank on all function card units is the same. Filter banks are used differently on single-layer function cards. There are two reasons behind this structure. First, local value groups in matrix data, such as pictures, are generally tightly linked, producing unique local patterns that are clearly recognized. Second, picture statistics and other signals at the local level are location-independent. In other words, if a motif can emerge in one area of a picture, it may appear everywhere, implying that units in various sections of the array have the same mass and recognize the same pattern. A function card's filtering action is mathematically referred to as a discretion, thus its name.

While the convolution layer's job is to discover local relationships between functions from the preceding layer, the pooling layer's job is to combine semantically related functions into one. Because the relative locations of the features that make up a feature may change somewhat, the position may be safely calculated by assuming that each feature's position is consistent. [56]On a single feature map (or a few feature maps), a typical pooling device calculates the maximum number of local patch devices, then shifts to smaller shifts and distortions. Convolution, non-linearity, and pooling are organized in two or three levels, followed by more convective layers and completely linked layers. It's simple to backpropagate the gradients through a single network over a typical deep network, allowing all of the filter banks' loads to be trained.

Deep and neural devices with multiple signals are those that have a higher level of power,

which allows for a higher level of sensitivity by counting as lower levels. In the paintings, the local connections between the edges create images, motifs clustered in space, and other areas. Similar situations exist in speech and the sound of phonemes, syllables, phones, words and phrases. The representation of the pooling varies slightly when the previous layer is not the same in position and shape.

ConvNets' [39]convolutional and pooling layers are based on the traditional notion of simple cells as well as the hidden neurobiology of complex cells. The overall architecture is similar to one of the visual cortex's LGN-V1-V2-V4-IT-level ventricular circuits. When the identical picture is shown to the ConvNet model and the monkey, the high-level unit activation in ConvNet explains half of the variation in the monkey's lower cortex's random group of 160 neurons. ConvNets originated from new cognitive machines, and their architecture is somewhat similar, but they lack a supplementary supervised learning algorithm that ends with backpropagation. Use the original ConvNet 1D called Delayed Neural Network to recognize phonemes and simple words.

Finding, segmenting, and identifying objects and areas in pictures has been a huge accomplishment for ConvNets. Traffic signal recognition, notably biological picture segmentation for connectedness, and the detection of human faces, text, pedestrians, and bodies in natural photographs were all tasks with a lot of labelled data. Face recognition has been ConvNets' most recent achievement.

It is critical that pictures be categorised at the pixel level, as this will be required for technological applications such as autonomous mobile robots and self-driving automobiles. ConvNet-based approaches are being used by companies like Mobileye and NVIDIA in their future car vision systems. Understanding natural language and voice recognition are two more applications that are gaining traction.

Despite these accomplishments, the visual vision and machine learning communities generally ignored ConvNets until the 2012 ImageNet competition. Deep convolution networks have produced amazing results when applied to a dataset of almost a million pictures from the web with 1,000 distinct classifications, substantially reducing the error rates of the best competing techniques. GPUs, ReLUs, a novel editing method called dropout, and ways to produce more training examples by deforming existing ones have all contributed to this achievement. ConvNets are now the dominating method to nearly all recognition and recognition problems, with certain tasks nearing human performance as a result of this success. For captioning, a new presentation combines ConvNets with iterative network modules in an amazing demonstration.

ConvNet's most recent design has 10 to 20 RLU layers, hundreds of millions of weights, and billions of connections. While training such a huge network took two years two years ago, improvements in hardware, software, and algorithms have cut training time to just a few hours. Because of the success of ConvNet-based vision systems, most major technology firms, such as Google, Facebook, Microsoft, IBM, Yahoo, Twitter, and Adobe, as well as a growing number of startups, are embarking on R&D projects and using ConvNet-based vision systems in goods and services. ConvNet is easily adaptable to chip or field programmable gate array hardware implementations. ConvNet chips are being developed by companies like as Nvidia, Mobileye, Intel, Qualcomm, and Samsung to allow real-time vision applications in smartphones, cameras, robotics, and autonomous cars.

So basically, CNNs process pictures as volumes, accepting a shaded image as a rectangular box where the width and height are measured by the quantity of pixels related with each measurement, and the profundity is three layers profound for each shading (RGB). These layers are called channels. Inside every pixel of the picture, the force of the R, G, or B is communicated by a number. That number is a piece of three, stacked two-dimensional grids

that make up the picture volume and structure the underlying information that is taken care of to into the convolutional arrange. The system at that point starts to channel the picture by gathering squares of pixels and searching for designs, performing what is known as a convolution. This procedure of example investigation is the establishment of CNN capacities.

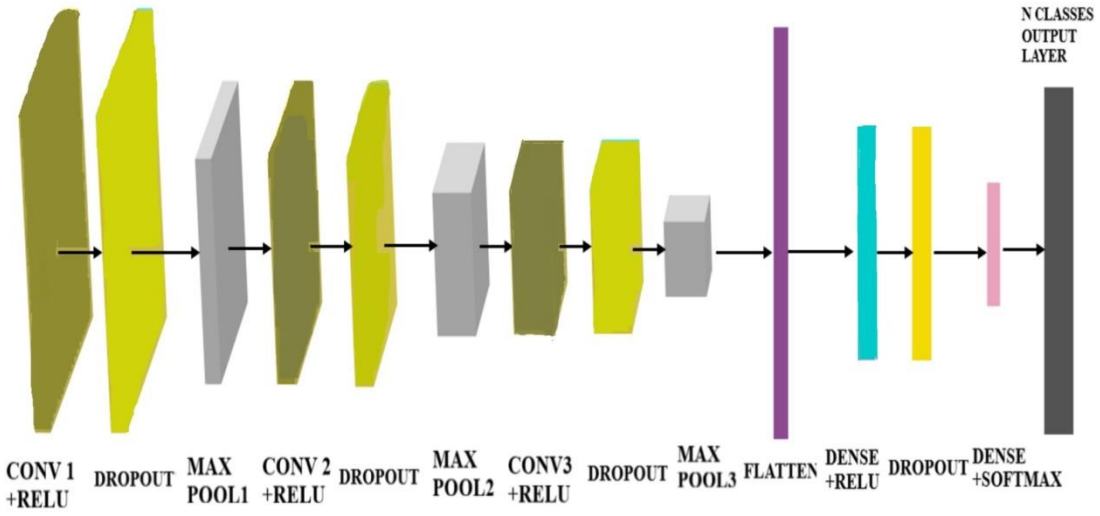


FIGURE 6: CONVOLUTION NEURAL NETWORK ARCHITECTURE

3.3 TRANSFER LEARNING

Traditionally, model is trained by using a huge dataset. The dataset is passed though Convolutional Neural architecture. After passing the images through the hidden convolutional layers, then putting weights on the layers after next iterations for error correction, the trained model is evaluated over test dataset. No information is retained. Therefore, even for doing similar kind of task, every time a new model has to be trained. This is time-consuming. While

in Transfer learning, a model is pre-trained using an already large available dataset and then that model is further used for training a new model solving the problem of having less publicly available labeled data. The previous learned learning is being kept so that for similar tasks that information can be used.

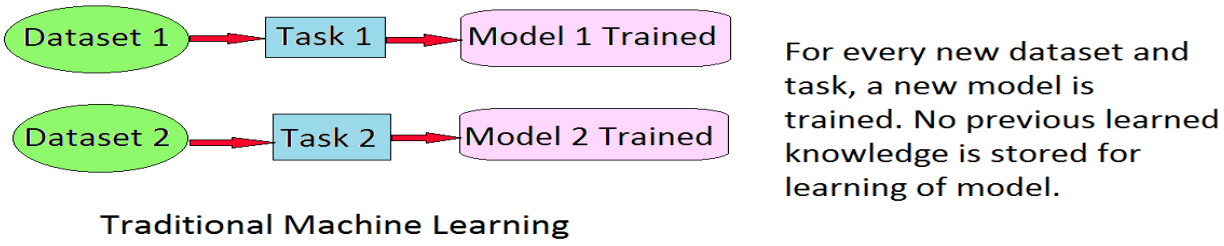


FIGURE 7: BASIC CNN

Formally defining, there is a big source domain $ḍ_s$ which contains a variety set of features F. For this the task is $Ţ_s$ and Marginal Probability Distribution is denoted as P(F), where F=(f_1, f_2, f_3, ……,f_n) all belonging to F. Now, there is a target domain Ɗ. It has a task $Ţ_d$ which is specified by a labeled set ¥ and a predictive function $\hat{g}$(.). $\hat{g}$ is learned by the training data. Its probability distribution is denoted as P(¥). Transfer Learning targets to increase the probability distribution P(¥|F) in the domain Ɗ. Logically, in this learning either, source domain $ḍ_s$ is not equal to Ɗ or source task $Ţ_s$ is not equal to target $Ţ_d$. The algorithm passes on the knowledge to the target model.

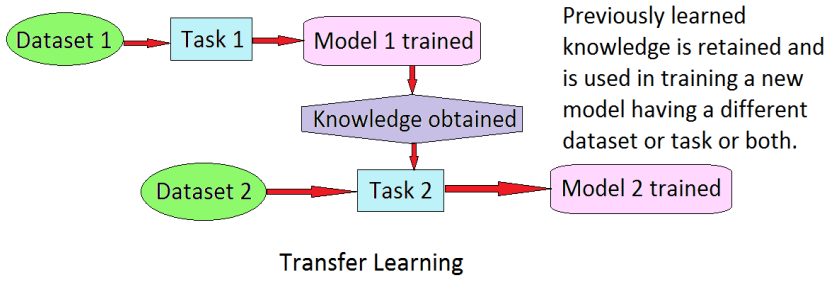


FIGURE 8: TRANSFER LEARNING

There are many pre-trained architectures. These are AlexNet, VGG, GoogleNet, ResNeXt, SqueezeNet, ResNet, DenseNet, ShuflleNet V2, MobileNet V2. So far, AlexNet, VGG, GoogleNet, ResNet have been used a lot for transfer learning to detect diabetic retinopathy. All the architectures are in layered form. At each layer, a different feature is extracted. The idea behind is to utilize the outputs all the layers of pre-trained models for training the target model except the last layer. The current model created will be shallow as no weights would have been applied. The target domain 'D's training dataset will be used to extract the features without updating the weights of the layers in the target model. This means that features are from the current dataset but the knowledge to how to extract those features and understand it are from the pre-trained network. Then finally, a classifier is applied for classification.

3.4 VGG MODEL

A 224 × 224 RGB picture is used as the input in the convNet-based VNG. During ImageNet training, the preliminary goal is to use RGB pictures with pixel values in the range of 0-255 and lower the average image output. Images of training are sent to decision-making groups. The VGG16 design has a total of 13 solutions and three connections. [32] VGG uses a smaller (3 * 3) filter with more depth than a bigger filter. Finally, the impact is the same as if you just had a 7 x 7 solution. VGG's design was based on a 224x224 pixel RGB image. In the

case of ImageNet rivalry, the creators trimmed out the inside 224x224 fix in each picture to keep the info picture size steady. The VGG convolutional layers utilize a little open field (3x3, the littlest conceivable size that despite everything catches left/right and up/down). There are additionally 1x1 convolution channels which go about as a straight change of the info, which is trailed by a ReLU unit. The convolution stride is fixed to one pixel with the goal that the spatial goals are saved after convolution. VGG is made up of three layers, each with 4096 channels: the first two levels have 4096 channels each, while the third layer contains 1000 channels, one for each class. VGG's hidden layers are all available to ReLU (a colossal advancement from AlexNet which narrows down preparing time). VggNet is designed to increase the number of system levels with 16 or 19 layers (VggNet-vd16 and VggNet-vd-19, respectively). On each convolutional layer, 3x3 size convolution channels with a step of size one are employed to reduce the number of weight parameters. Vgg-f, Vgg-m, and Vgg-s are three CNN architectures proposed in article [10] to understand how different CNNs are compared to one another and to existing best in class shallow representations. Five convolutional layers, three max-pooling layers, and three fully associated layers make up CNN models. These, in any case, have different channels, pooling sizes, strides, and open field capacity.

Another VGGNet variation contains 19 weighted layers, 3 fully connected layers, and 16 convolutional layers, as well as the same 5 pooling layers. VGGNet has two completely connected layers, each with 4096 channels, followed by another fully connected layer with 1000 channels, which predicts 1000 tags in both variations. The soft max layer is used for classification in the final fully linked layer.

Method of design: the first two layers are the solution process with a 3 * 3 filter, and the first two layers use a 64 filter which generates a 224 * 224 * 64 code based on the configuration used. The filter is always 3 * 3 with a sequence of 1. After that, the filter layer can be used with the latest max * 2 * 2 size and 2 outputs which reduces the level of 224 * 224 * 64 to ' of 112 *

112 * 64. It is based on two existing 128-filter systems. As a result, the size is now 112 * 112 * 128. The volume is decreased to 56 * 56 * 128 once the layer quantity has been used up. As a result, several layers are added to the sample layer, each with 256 filters, reducing the size to 28 * 28 * 256. The multi-layer filtering of 3 layers can be separated from the max-filter layer. problem of class 1000.

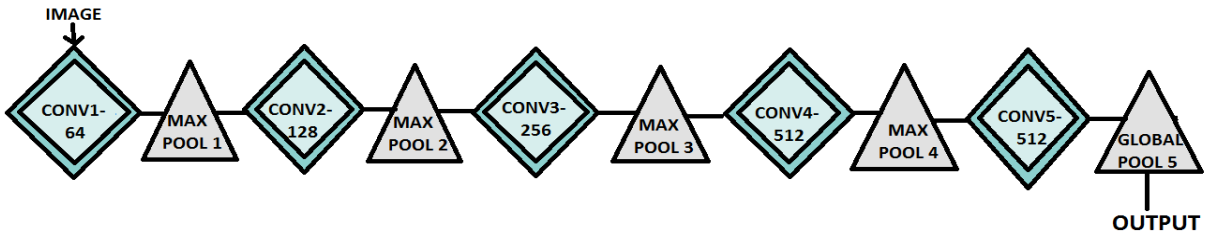


FIGURE 9: VGG NETWORK ARCHITECTURE

3.5. U-net

U- Net is introduced in 2015 [12], especially for processing biomedical images. Its basic foundation is upon CNN's conventional architecture. The general convolution neural network is programmed to classify images, where the image is the input and the output is the desired label. U-Net helps to distinguish images pixel by pixel and apply classification over it. It locates the areas of abnormalities. Input and output share the same sizes. U- Net is a kind of architecture that is in U shape. It has two parts that are the left and right sides. The left side, which is called the contracting path, is the usual convolution neural network. The ride side, which is called expansive path, is there so that up-sampling can be done. It is a stack of deconvolution blocks. Since, it extracts semantic feature, therefore, it is able to do better image classification. The issue in U-net is that the input compress, which creates a bottleneck where all features are not

passed into the decoder.

The usage of different network combinations is determined by the job distribution, which implies that pictures are assigned to a single character class. However, the product must be incorporated in many visual operations, particularly in biomedical imaging, where class names must be allocated to each pixel. Thousands of training pictures are also often applied to biological activities. As a result, a research is carried out for the network in the sliding window to predict the class label for each pixel using the local area (region) around the pixel as input. To begin, this network may be searched in the surrounding region. Second, the number of correction points presented is larger than the number of training photos. By a considerable margin, the outcome wins the European Championship competition in the ISBI 2012 tournament.

However, it has two flaws. For starters, it's sluggish since each patch requires running the network individually, and there's a lot of redundancy due to patch duplication. Second, there is a misalignment between localization accuracy and context usage. Larger patches need additional accumulation layers, lowering localization accuracy, but smaller patches only enable the network to perceive a limited amount of information. A more recent technique proposes a classifier output that considers several layers' characteristics. Good localization while also utilising good context.

A more complex architecture was created, known as a "fully connected network." The authors adapted and scaled this architecture to work with minimal training images and create more accurate segmentation. The basic concept is to successfully strengthen the usual contracting network layer by replacing the bundle operator with an upgrade operator. As a result, these layers improve the output resolution. The contract path's high definition functionality is combined with the improved output for localization. On the basis of this data, successive layers

of confidence can learn to acquire more accurate findings.

The presence of multiple service channels [41] at the top of the structure is a significant change, as it allows the network to send contextual information in high-definition columns. As a result, the wide path via transmission is more or less symmetrical, and it gives an organised framework. There are just pixels in the partition map, which represents the entire environment as seen in the insertion image, because the network does not have completely integrated layers and only utilises the valid part of each fidelity. The AnoverLap-Tile method allows you to arbitrarily split big pictures using this strategy. The missing context is produced by reflecting the input picture and estimating the size of the image's border region. If you want to use the Internet for huge photos, you'll need to use this tiling approach; otherwise, GPU memory will determine the resolution.

Excessive data augmentation is achieved by deforming the available training pictures with elastic deformation. This allows the network to learn the invariance of such aberrations without having to look at them in a corpus of annotated images. [34]This is particularly significant in biomedical segmentation since deformation was previously the most prevalent change in tissues, and it can successfully imitate actual deformation. Separation of contact objects of the same type is another problem in many cell segmentation tasks. To achieve this, we suggest employing a weighted loss, in which the separated background labels between the contact cells are given a high weight in the loss function.

U- Net is a kind of architecture which is in U shape. It has two parts that are the left and right sides. The left side, which is called the contracting path, is the usual convolution neural network. A total of two convolutional blocks are used. Each block consists of two 3 X 3 unpadded convolutional blocks, each of which is followed by a ReLU (Rectified Linear Unit) activation function and a 2 X 2 max pooling layer with stride equal to two, allowing for

downsampling. The number of channels of features is doubled after each downsampling. The ride side, which is called expansive path, is there so that up-sampling can be done for the features by adding 2 X 2 up convolution. The number of feature channels is cut in half in this step. A concatenation layer is added to this phase, along with feature maps from each contracting path and two 3 X 3 convolution layers, each of which is followed by a ReLU (Rectified Linear Unit) activation function. Cropping is required because the border pixels are lost after each convolution stage. Finally, every 64 component vector of features is matched to proper or accurate classes using a 1 X 1 convolution layer.

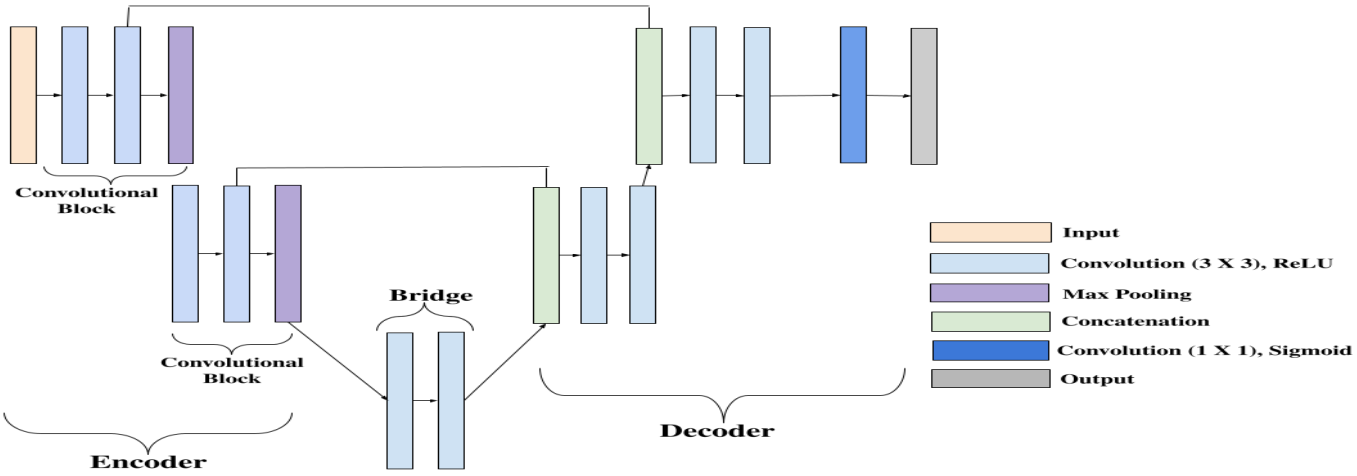


FIGURE 10: U-NET NETWORK ARCHITECTURE

3.6. Residual Unit

Stacked residual units are applied successively in residual networks [33]. Every unit is expressed in the following form:

$$k = y(x) + f(x, w) \qquad (1)$$

$$xi = A(k) \qquad (2)$$

where x is the input, xi is the output of the I-th residual unit, f(.) denotes residual function, A(k) denotes the activation function, y(x) gives the identity mapping. Numerous fusion of convolutional layers, ReLu activation function and batch normalization are present in a single residual unit.

Consider H (x) as a low-level mapping appropriate for a number of stacked layers (not necessarily the full grid), with x being the first of these layers' input. [58]Assume that many nonlinear layers can asymptotically approach the complex function 2, which corresponds to the assumption that they can asymptotically approximate the other functions, i.e. H (x)-x (assuming that the input and output have the same dimensions). We must deliberately make the stacked layers resemble the residual function F (x): = H (x)-x, rather than expecting them to be near to H (x). As a result, the original function becomes F (x) + x. As a result, the original function becomes F (x) + x. Both must be able to approach the target function asymptotically (for example, hypothesis), although the easy-to-learn approaches may differ.

The counter-intuitive phenomena of deterioration is the rationale for this restatement. If the additional layer can be built as an identity map, the deeper model's training error should not be higher than its shallower version. The degradation issue suggests that the solution may have trouble reaching the identity assignment via several non-linear layers. [35]If the identity mapping is optimum for the remaining learning reconstruction, the solver can simply set the weights of many nonlinear layers to zero to approach the identity mapping.

Identity mapping is unlikely to be ideal in practise, but the reformulation can aid in preconditioning the problem. If the optimum function is closer to an identity mapping than to a null mapping, finding perturbations with reference to an identity mapping should be easier for the solver than learning the function as a new one. Experiments demonstrate that the residual

functions learnt in general have modest responses, implying that identity mappings are a good preconditioning method.

Eqn (1) generates quick connections with no extra parameters or computational complexity. When comparing regular and residual networks, this is not only appealing, but also crucial. At the same time, we may compare residual and regular networks that have the same number of parameters, depth, width, and computation cost (except for a slight elemental increase).

The equation requires that the dimensions X and F be equal (1). If this isn't the case (for example, to modify the input/output channels), we can make a linear projection Ws to fit the dimensions using linear connections:

$$y = F(x, \{Wi\}) + Wsx. \qquad (3)$$

We can also use the square matrix Ws in the equation. (3). However, we seem to experiment that identity mapping is sufficient to solve the decomposition problem and is economical, and therefore Ws is used only when the dimensions are fitted.

The remaining function F's form can be changed. A function F with two or three layers is included in this paper's experiments, however additional levels are feasible. However, if F only has one layer, equation (1) is the same as the linear layer: $y = W1x + x$, with no benefits.
It should also be noted that, [29]while the above notation refers to linked layers for ease of understanding, it also applies to convolutional layers. Several convolutional layers can be represented by the function F (x, Wi). On two function maps, do element-wise addition channel by channel.

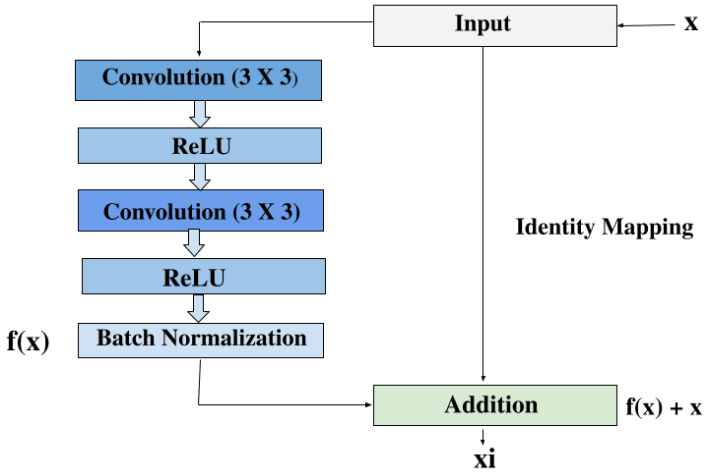


FIGURE 11: RESIDUAL UNIT

3.6.1 BATCH NORMALIZATION LAYERS IN A BUILDING BLOCK

A Batch Normalization layer's primary function is to normalize activations for faster convergence and better performance. A single residual unit's capacity can be increased by using the BN layer. The following equation is used by a BN layer to execute an affine transformation:

$$y = \gamma x + \beta, \qquad (4)$$

For each activation in the function graph, and are learnt. Experiments have revealed that the learnt and can be very close to zero. This indicates that if the learnt y and are near to 0, the activation associated with them is regarded worthless. [58]Deep weighted ResNets, which have observable weights at the end of their construction blocks, learnt to evaluate if a particular residual unit is helpful in a similar way. As a result, the BN layer at the end of each residual unit is a generic form, allowing for the determination of whether or not each residual unit is

helpful. As a result, the degree of freedom gained by linking the BN layer's and can increase the network's capabilities.

3.6.2 ReLUs IN A BUILDING BLOCK

It's critical to incorporate ReLU in the remaining units' building blocks for nonlinearity. However, we discovered that performance varies depending on the number of ReLUs and their placement. This may be compared to the original [59]ResNet, and the findings indicate that the performance improves as the network deepens. However, if the depth is greater than 1,000 layers, overfitting occurs, and the output is not as precise as the smaller ResNet..

First, it is found that using ReLU after adding residual units can adversely affect performance: x l k = ReLU (F(k,l)(x l k−1) + x l k−1), (5) ReLU appears to include unfiltered non-negative components. After each addition, it was discovered that removing ReLU from the original [59] ResNet through a quick switch improves performance somewhat. This can be understood as ReLUs providing a non-negative input for the subsequent residual unit after the addition, so the quick link is always non-negative, and the convolution expert is responsible for generating a negative output before the addition, reducing the overall network architecture capacity. The ResNets that have been suggested have already been activated. Pre-activated residual units, which set BN layers and ReLU before (rather than after) the convolutional layer, also solve this problem.

$$x\,l\,k = F(k,l)(x\,l\,k-1) + x\,l\,k-1, \quad (6)$$

To build an identity route, ReLU is deleted after insertion. As a result, even at depths of over 1,000 floors, overall performance improves considerably without the need for extra installation. A weighted residual network design is also suggested, which inserts ReLU in the residual unit (rather than locating ReLU after adding) to establish an identity path, although this structure is

not suited for depths higher than 1000 layers.

Second, it is discovered that using a large number of ReLUs in each remaining unit's blocks might have a detrimental impact on performance. It was discovered that removing the first ReLU from the blocks of the first remaining unit improved performance over the blocks. To assure nonlinearity, we discovered that removing the first ReLU from the array is preferable, while keeping the second ReLU. The blocks are converted to BN-ReLU-conv-BN-conv [60]by deleting the second ReLU, and it is apparent that these blocks are convolutional layers without ReLU to mutually degrade the representational capabilities. The blocks are changed to BN-conv-BN-ReLU-conv if the first RELU is removed, in which case the two convolutional layers are separated by a second ReLU, ensuring nonlinearity. As a result, the remaining ReLUs can be eliminated to enhance network performance if a suitable number of ReLUs are employed to assure nonlinearity of space deterioration.

3.7. RESIDUAL U-NET

Since Residual U-net forwards the input layer information, it is referred to as weighted Res-Unet in many research papers [13]. It is a combination of U-net and residual neural network. It provides two advantages. First, it helps in reducing the computations of training. Second, it propagates information without degrading the model between low and high levels. Both of the advantages are achieved with fewer parameters. Meanwhile, it attains the preferable quality of performance.

In this experiment, five layer of deep residual U-Net [59]is applied. The neural network contains encoding part, bridge part and decoder part. The encoder encodes the input image into a condense form, while the decoder again decodes it into the pixel wise formation of the input and puts into required classification which is basically semantic segmentation [35]. The bridge

part joins the encoder and decoder. All these parts are made up of residual units. The input and output of every residual unit is tied up by an identity mapping.

Encoding part has two residual units. Each residual unit is made up of two convolutional layers with 3 X 3 kernel size and ReLU activation function, a batch normalization layer, and a dropout layer. After this, addition is done with the input layer, calling it Layered Output. Then, 2 X 2 max pooling layer with stride equal to two is applied, so that down-sampling can be done. The output is passed onto the next block. Bridge part has a single 3 X 3 convolutional block with two convolutional layers. The decoding part again comprise of two convolutional blocks. In every block there is convolutional transpose of input layer which is added with the corresponding Layered Output of encoding layer, calling it Decoded Output. It is followed by two convolutional layer having 3 X 3 kernel size and RelU activation function, a batch normalization layer. The obtained output from the block is again concatenated with Decoded Output. At the last, a 1 X 1 convolutional layer and a sigmoid function is utilized so that multi-channel feature maps can be projected giving the required segmentation results.

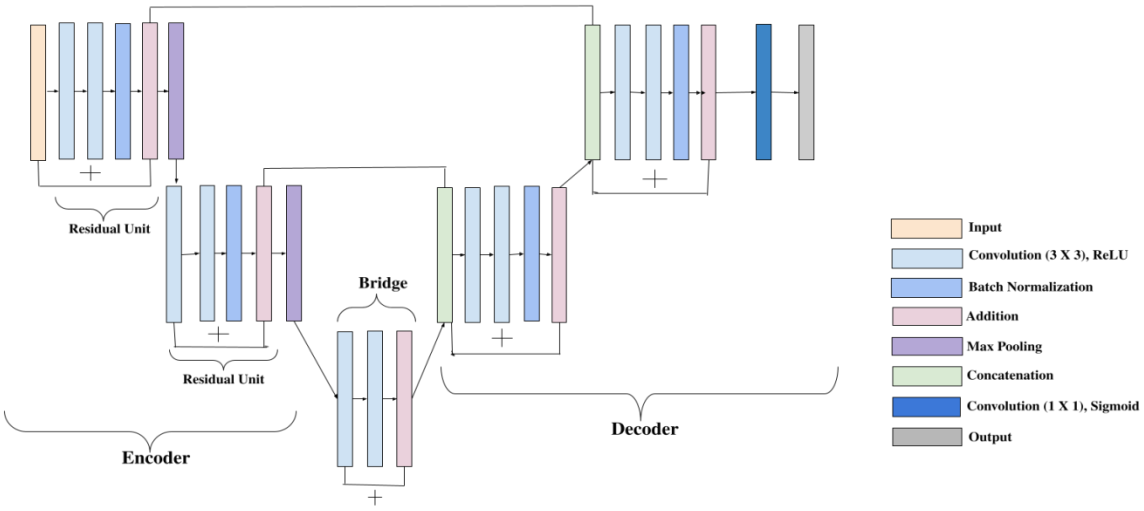


FIGURE 12: RESIDUAL U-NET NETWORK ARCHITECTURE

3.8. Y-NET

Y-net [14] is also a three-part procedure, just like U-net. It has three paths: encoding, bridge, and decoding. It uses two encoders for instance-level segmentation. The input layer for both encoders is the same and the output of two convolutional blocks from the first and second encoders are added and then transferred to the corresponding stage of the decoder block. It helps to keep together features obtained in the condensed form in a much secure way. The decoding goes the same as the U-net decoding block. Convolutional transpose of the input layer is done. It is then added to the previously obtained added result of the corresponding outputs from the two encoders' convolutional blocks.

Different convolutional blocks can be used at any level without affecting the architecture of the network. It proves to be flexible and more combinations of parameters can be applied according to the requirement of the dataset and better performance of the model.

In this model, two encoding blocks passes through one bridge into one decoding block. Every encoding block has two convolutional blocks. Each convolutional block comprise two 3 X 3 convolutional layer each with a ReLU activation function and a dropout layer followed by 2 X 2 max pooling layer with stride equal to two so that down-sampling can be done. The output is then passed into the next block as an input layer. The output of every convolutional block of the first encoding block (before pooling layer) is concatenated with the output coming out from the corresponding convolutional block of the second encoding block, calling this result as Block Encoded Output. It helps to keep together features obtained in condense form in much secured form. This is done after every convolutional block.

Input that goes inside the bridge is the addition of the output (after pooling layer) of the last

convolutional blocks from the two encoding blocks. It has two 3 X 3 convolutional layers of filter size 128 with ReLU activation layer. The decoding goes same as the U-net decoding block. Convolutional transpose of the input layer is done. It is then concatenated with the previously obtained added result of the corresponding output from the convolutional blocks of the two encoders. It is followed by a dropout layer and two 3 X 3 convolutional layer. The output is then passed into the next block in the decoder. Lastly, multi-channel feature maps are obtained by a 1 X 1 convolutional layer and a sigmoid function so that optimal segmentation results can be achieved.

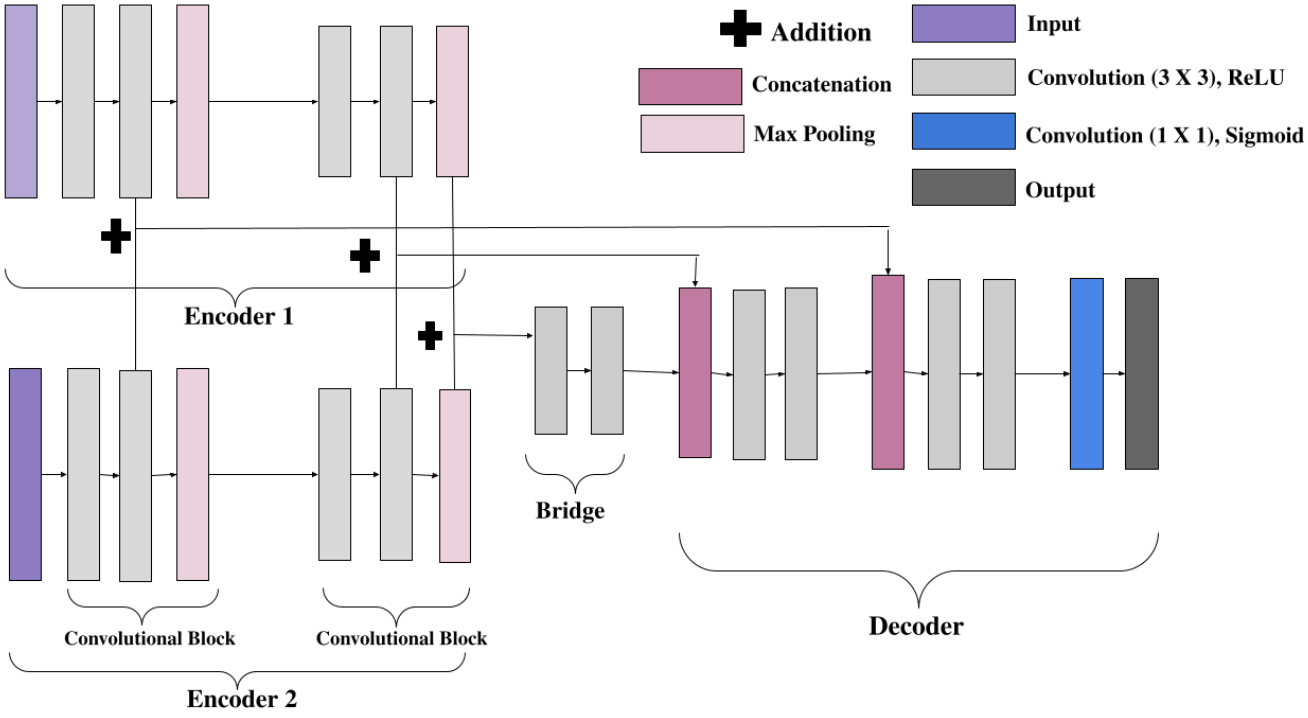


FIGURE 13: Y-NET NETWORK ARCHITECTURE

How Y-net is different from U-net:

In U-net, same convolutional block output which is passed into the next blocks goes into the corresponding decoding blocks. While in Y-net, first the output of two parallelly present convolutional blocks from first and second encoder is added and then passed into the corresponding level of decoder block. Here, it is not necessary to have the same spatial value at every point [3]. The encoding and decoding convolutional blocks are therefore, general in nature. Different convolutional blocks can be used in any level without affecting the architecture of the network. It proves to be flexible in nature and more number of combinations of parameters can be applied according to the requirement of dataset and better performance of model.

3.9. IMAGE AUGMENTATION

Data Augmentation [60] helps to increase the amount and diverseness of data during the scarcity of publicly available data. In image augmentation, the raw input images are rotated, sheared, clipped, zoomed, flipped, contrast is changed. In this way, different views of the unexpected cases of diabetic retinopathy can be obtained. Many techniques and APIs are accessible for augmentation. Histogram Equalization is used for adjusting the contrast of images. The raw input images obtained do not always have similar brightness. So to have a uniform luminosity, it spreads the most efficient pixel to all over the image. The APIs have Image Data Generator, in which various parameters are passed. In Image Data Generator, the rotation range, shearing range, etc. can be defined and passed according to the requirement.

DATA AUGUMENTATION BASED ON BASIC IMAGE MANIPULATIONS

Geometric Transformations

[61]This section covers a variety of magnification techniques based on geometric changes, as well as a variety of other image processing aspects. The magnification class described below is distinguished by the simplicity with which it may be used. Understanding these changes lays a solid foundation for future data augmentation research.

The chance that the label will be retained after transformation is referred to as the security of a data augmentation approach. Spins and flips, for example, are typically safe for ImageNet problems like cat vs. dog, but not for Caller ID jobs like 6 vs. 9. The model's ability to generate a response indicating that it is not certain would be harmed if the transformation did not preserve the tags. is a forecast he made. However, following augmentation, fine-tuned labels are necessary to do this. The model might acquire more robust confidence predictions if the image's label after transformation without label preservation is about [0.5 0.5]. Creating complex labels, on the other hand, is a difficult task.

Given the challenge of creating refined tags for post-aggregation data, it is important to consider the "security" of magnification. It depends a bit on the domain, challenging the development of a generalized aggregation policy. There is no image transformation function that does not change the label to the point of certain transformation distortions. This shows the specific conception of magnification data and the challenge of developing a generalized magnification policy. This is an important consideration for geometric magnifications.

Flipping

The reversal [61] of the horizontal axis is far more common than the reverse of the vertical axis. This is one of the simplest methods, and it has been proved in datasets such as CIFAR-10 and ImageNet. Text-recognition data sets, such as MNIST or SVHN, are not tag-

preserving modifications.

A rotated image, or inverted image, is a more formal term, a static or moving image created by returning a mirror to the horizontal axis of an original (a curved image is reflected on the vertical axis). . will be the picture. Many printer manufacturers have developed the ability to return images while making print boards, but many, especially the early ones, have images that can be reversed.

Color space

A dimension tensor [61] (height, width, and color channels) is commonly used to encode digital picture data. Expanding the color channel space is another extremely useful approach to use. Isolating a color channel, such as R, G, or B, for very simple color enlargements is useful. By showing this matrix and adding two zero matrices of additional color channels, a picture may be instantly transformed to a color channel. Furthermore, RGB values may be readily adjusted using basic array operations to raise or reduce the image's brightness. The most advanced color magnifications are obtained by constructing a color histogram of the picture. Changes in brightness are caused by changing the intensity value of these histograms.

Cropping

Picture cropping [61]can be used as a processing step for image data with mixed height and breadth by cropping the centre of each image. Furthermore, random cropping may be utilised to create effects that are similar to translation. Random cropping differs from rotation in that cropping decreases the size of the input, such as (256 256) (224, 224), whereas translation retains the image's spatial dimension. This may or may not be a reserved conversion, depending on the deduction threshold used for delimitation.

Rotation

Rotate [61] the picture right or left on an axis between 1 and 359 degrees to make rotation increments. The rotation degree parameter has a big influence on the rotation increments' safety. Slight rotations between 1 and 20 or - 1 to - 20 may be beneficial for digital recognition tasks like MNIST, but when the degree of rotation rises, the data label is lost following transformation.

Translation

To produce rotation increments, rotate the image right or left on an axis between 1 and 359 degrees. The rotation increments' safety is heavily influenced by the rotation degree parameter. For digital recognition tasks like MNIST, little rotations between 1 and 20 or - 1 to - 20 may be advantageous, but as the degree of rotation increases, the data label is lost during transformation.

Geometric transformations are an excellent way to deal with positional biases in training data. There are numerous types of bias that might cause the distribution of training data to differ from the distribution of test data. Geometric transformations are a useful option if positional biases are present, such as in a facial recognition dataset when each face is properly centred in a frame. Geometric transformations are helpful not just because of their tremendous capacity to overcome positional biases, but also because they are simple to utilise. To begin with, there are numerous image transformation libraries that do operations like flipping and painless rotation. Increased memory, computational transformation costs, and additional training time are some of the drawbacks of geometric transformations. To guarantee that geometric modifications such as translation or random cropping do not modify the image's label, some geometric transformations must be tracked manually. Finally, deviations of training data from test data are

more complicated than position and translation deviations in many covered application areas, such as medical image analysis. As a result, the places and times when geometric transformations can be employed are limited.

Color Space transformations

The picture data is divided [61] into three stacked matrices, each with a length and width of length and width. The pixel values of a single RGB colour value are represented by these matrices. The most prevalent provoking image recognition issues are minor biases. As a result, the notion of colour space alterations, often referred to as photometric modifications, is very simple to grasp. Going through the photos and reducing or increasing the pixel values by a standard value is a simple fix for images that are excessively bright or dark. Merging separate RGB colour models is another quick colour space modification. Limiting pixel values to a specified minimum or maximum value is another modification. The natural color representation of digital images is suitable for many enhancement strategies.

Image editing applications may also be used to convert colour spaces. A colour histogram is made up of the pixel values of each RGB colour channel picture. This histogram may be used to add filters to the picture that alter the colour space characteristics. Adding colour space allows you a lot of creative freedom. Changing the colour distribution of photographs can be a fantastic way to overcome the lighting issues that data testing faces.

Color space conversions have the same drawbacks as geometric conversions in terms of memory use, conversion costs, and training time. Color transformations also have the potential to erase essential color information, thus they are not necessarily label-preserving transformations. When the pixel values of a picture are reduced to mimic a darker environment, for example, no objects in the image may be seen. Image analysis is another indirect example

of maintaining color changes on a label. CNNs are used in this application to visually forecast an image's emotion score, such as extremely negative, negative, neutral, positive, or very positive. The presence of blood is one of the markers of a bad/very unfavorable picture. Blood's dark red hue plays a vital role in differentiating it from water or other colors. If the color space is continuously changed, the model will do badly in imagery analysis because it will be unable to distinguish red blood from green color. Color space transformations, in effect, replace color aberrations in the data set with spatial characteristics. Color, on the other hand, is a key differentiating characteristic for various activities.

3.10. FINE TUNING

Fine Tuning [55]is an approach to use an already trained model for doing a relatable task. This is a state-of-art where the steps are taken slowly so that adjustments can be made to the weights at the final stage. At the point when you have little datasets (for example not many 1000s).When the dataset used to prepare the pre-prepared model is fundamentally the same as or equivalent to the new dataset.

A collection of randomly formed weights is used to initiate the repeated weight update. Prior to the start of the training phase, CNN initializes the weights in each folding layer, using values selected at random from a normal distribution with a mean of zero and a modest standard deviation. Given CNN's high number of weights and the scarcity of tagged data, iterative weight updating with random weight initialization may result in an undesirable local minimum cost function. Alternatively, the convolution layers' weights can be set to the weight of a pre-trained CNN with the same architecture. A huge collection of tagged data from another application is used to construct a pre-trained network. CNN's pre-trained weights are known as fine-tuning, and they've been effectively employed in a variety of applications.

The fine-tuning process begins by copying (transferring) the weight we wish to train from a network that has already been created. The last completely linked layer, whose number of nodes is determined by the number of classes in the data set, is an exception. It is standard practise to replace the pre-prepared CNN's final completely connected layer with a new fully connected layer with the same number of neurons as the number of classes in a new target programme. We are working on 2nd and 3rd grade categorization problems in our research. As a result, depending on the application findings, the newly connected layer will have two or three neurons. Once the last layer's weight is entirely connected, a new network may be created in the layer template layer, starting with the last layer's perfection and then changing all layers on CNN.

Consider the L-layer CNN, which combines the last three layers. Let al represent the network's l-th learning speed. One may fine-tune the parameters by not putting al = 0 for l not equal to L. We may fine-tune it by just specifying the last (new) network layer. The progress of linear classification using features produced in the L-1 layer is linked to this degree of accuracy. Similarly, by setting al = 0 for l L, L - 1, the last two network layers may be fine-tuned. This rule corresponds to one hidden layer, which may be thought of as an artificial neural network. L is an indirect categorization that employs two levels of characteristics. In the same way, appropriately aligned L, L-1, and L-2 layers are essential. The pre-designed CNN will be more compatible with the manual application if prior solution layers are incorporated into the update process, although this may need proper account information to avoid compatibility.

In general, the early layers of a CNN learn low-level picture attributes that are important to most viewing activities, while the lengthy layers teach high-level image features that are useful to hand-held applications. As a result, learning the transfer generally just requires changing the last layers. However, if there is a large distance between the source and destination

applications, the initial layers may need to be modified as well. As a result, starting from the final layer and gradually increasing the increment until the desired performance is attained in the upgrading process is an efficient refining approach. The tuning of the last twisted layers is referred to as a "shallow depth correction," and we feel it is necessary. We'd like to stress that this network stands out as a source of features for the adaption of particular ideas that follow. It is distinct from the point at which the complete network fits at the same moment.

3.11. OPTIMIZER

Optimizers provide a better learning to the model. This improves the model perfo [59]rmance if appropriate optimizer is used according to the method and the size of the dataset. RMSProp, adam (adaptive moment estimation), sgd (stochastic gradient descent), adadelta are few well known optimizers used in classification models. One needs to decide the learning rate to successfully obtain a good result. Like in RMSProp and gradient descent, having low learning rate will improve the model performance in transfer learning models. Whereas there optimizer like adadelta which does not need any learning parameter. Adadelta eventually improves the CNN model performance.

The activation function is an important element of neural networks. ReLU is the most flexible, stable and used. Most commonly used in the 1980s, the Sigmoid function is only used today for certain types of output levels due to rating issues. Without them, deep learning models would behave just like linear functions. The most widely used activation function in the 1980s was the sigmoid, which is a continuously differentiating regular nonlinear function. Despite its interesting properties, it has a very important concern called the extinction gradient problem. The first Sigmoid variant arrives not far from the origin, which affects network loss optimization due to almost zero grading. ReLUs were used to develop limited Boltzmann (RBM) machines by converging with ReLUs to stepped sigmoid units. The performance of

Sigmoid, Tanh and ReLU is comparable and although Sigmoid is more biologically plausible, Tanh and ReLU have been found to be more suitable for use as an activation function for training multilevel sensors. ReLU networks generally perform better despite their non-zero differentiation and non-strict linearity. Additionally, ReLU networks lead to trivial representations, which is useful both because information is represented more strongly and leads to significant computational efficiency. In addition, the simplicity and derivative of the function reduces computation time, which is important when working with large networks. The fixed value of the rating allows for deeper meshes to be designed, helping to avoid the problem of loss of slope. ReLU has now become the default learning enable feature. Many other enabling features have been released, such as LeakyReLU, but they do not show significant performance gains.

CHAPTER 4
PROPOSED METHOD: RESIDUAL Y-NET

It is the combination of both Residual Unit and Y-net. The architecture is in the form of Y-net applying residual unit in the encoding and decoding block. It serves quality performance. At every level, the features are added from every other encoding level, therefore there is less degradation in the model. Just like Y-net, it has two encoders, one bridge and one decoder. Each encoder has two residual units. Every residual unit contains two convolutional layers with 3 X 3 kernel size and ReLU activation function, a batch normalization layer, and a dropout layer. After obtaining Layered Output, a 2 X 2 max-pooling layer with a stride equal to two is applied in the block for down-sampling. This Layered Output is added with its corresponding Layered Output from the second encoder to retain the best segmentation features and label in the best possible way, producing the Block Encoded Output. Just like in Residual U-net, it helps to keep features obtained in the condensed form. The difference is that this Block Encoded Output of Residual Y-net contains double number features. It increases the possibility of better label classification. The bridge part has two 3 X 3 convolutional layer in which the addition of the output from the two encoders is passed on as input. The decoding part is the same as Y-net. The Decoded Output is generated by concatenating the Convolutional Transpose of the input with Block Encoded Output from corresponding levels. It is followed by two convolutional layers having 3 X 3 kernel size and RelU activation function, a batch normalization layer. The obtained output from the block is again concatenated with Decoded Output of the current block.

The result of the last residual block of the decoder will be a multi-channel feature map and to perform classification over the feature map, a dense layer has been added. The size of the dense

layer output is equal to the number of classes in the dataset.

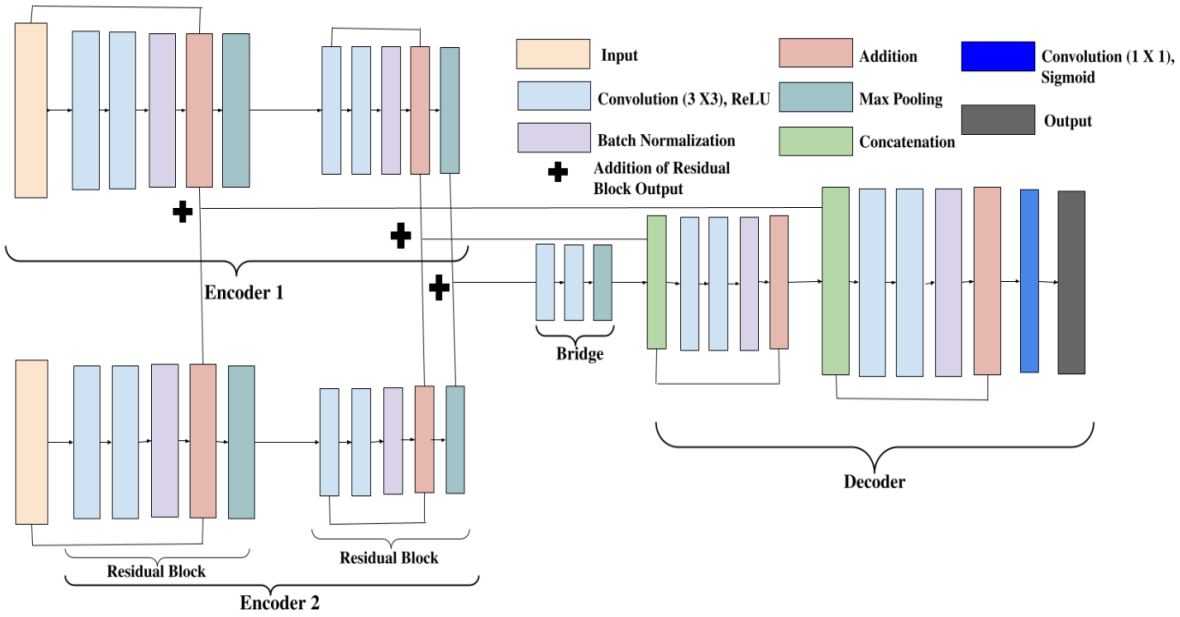


FIGURE 14: RESIDUAL Y-NET NETWORK ARCHITECTURE

How Residual Y-net and Residual U-net is different:

The shape of the network architecture is different. Residual Y-net is in Y-shape topology while residual U-net is in U-shape topology. In Residual U-net the Layered Output goes into the decoder, while in Residual Y-net two Layered Outputs is first combined and then goes into the decoder. More number of features can be retained due to having two encoders. Just like Y-net, it is flexible. Residual Y-net is not bound into spatial constraints. Therefore, different residual blocks can be used in any level without affecting the architecture of the network. The number of residual blocks, parameters and layers is decided depending upon the requirements.

CHAPTER 5

EXPERIMENT AND RESULTS

5.1. DATASET

The datasets are taken from APTOS 2019 blindness detection competition which took place on Kaggle. The datasets are classified into five categories, which are No_DR, Mild, Moderate, Severe and Proliferate DR. The total numbers of images are 13000. The competition dataset is unbalanced. From that dataset, Dataset1 which is unbalanced (Figure 15) and Dataset 2 which is a balanced dataset are derived to observe the performance of the proposed algorithm on both datasets and it consists of 3662 and 4918 images respectively. As in the bar graph can be seen (Figure 16), the number of pictures in each label is quite balanced, i.e. has a similar distribution of data among the classes, so the chances of random division into training and test set will bring better diversity.

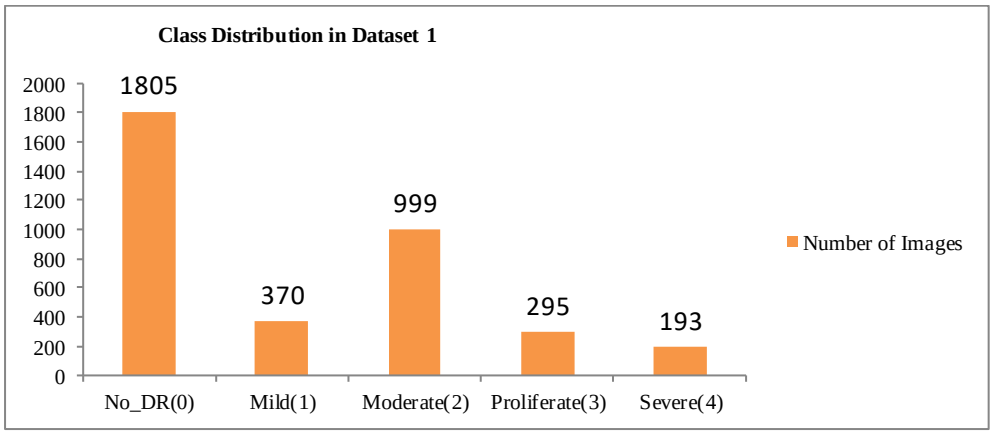


FIGURE 15: DATASET 1 DETAILS

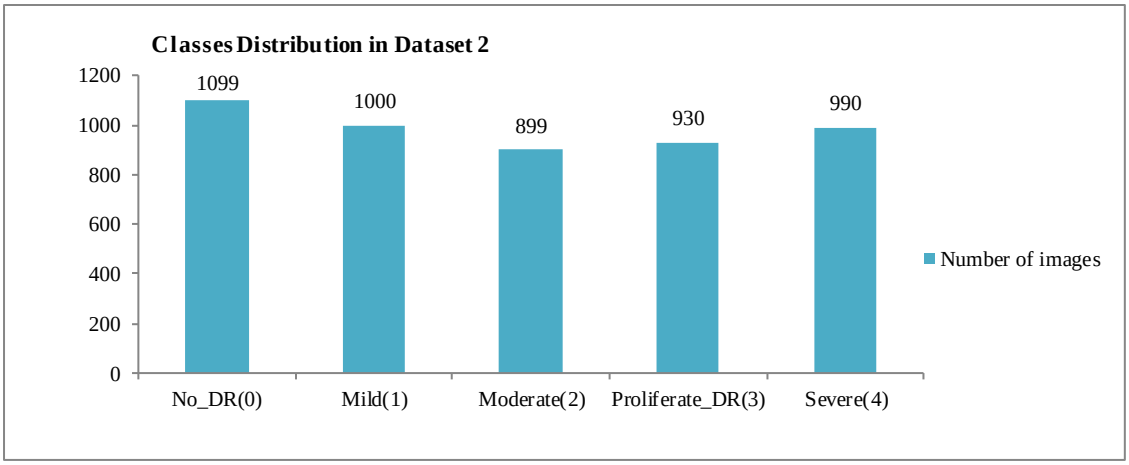


FIGURE 16: DATASET 2 DETAILS

5.2. Experimental Setup

The experiment executes on Google Colaboratory. The images are augmented. Image Augmentation increases the amount and diverseness of data during the scarcity of publicly available data. In image augmentation, the raw input images are rotated, sheared, clipped, zoomed, flipped, and contrast is changed. The APIs have an Image Data Generator, in which various parameters are put. The rotation range, shearing range, and other parameters can be specified and passed to Image Data Generator according to the requirements. Adam optimizer has been used in all the networks with categorical cross-entropy as a loss function. The training and testing data is split into 80-20 ratio randomly.

5.3. Results

The network's requirements decide the number of epochs. More blocks will result in the saturation on the image condensation since the size of the image is small. So basically, the

independent learning of individual neurons decreases. After adding the required number of dropouts, performance gets better. In proposed method which is Residual Y-net, residual units are added, which helped in increasing the high resolution information (segmented results). The parameters increased as compared to Y-net, U-net and Deep Residual U-net. But, there is significant decrease in training epochs. It takes 50 epochs to train on Dataset 1(Figure 17). The Training accuracy is 98.92%, and test accuracy is 90.39%. On Dataset 2, within nine epochs, the test accuracy reaches 93.60%, though the training accuracy is 98.65% (Figure 19).

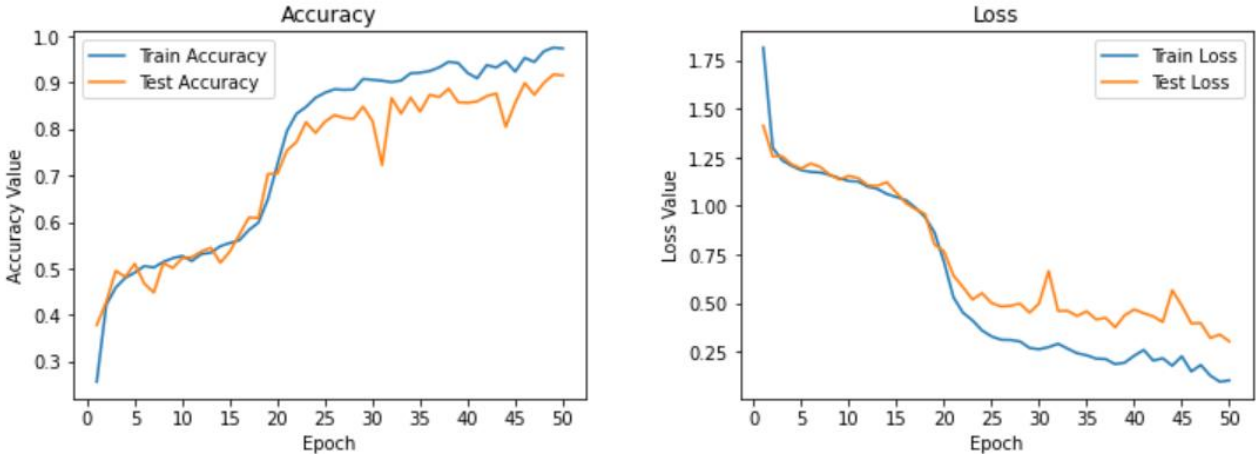


FIGURE 17: RESIDUAL Y-NET ACCURACY AND LOSS DATASET 1

Classification Report

	precision	recall	f1-score	support
0	0.92	0.97	0.94	219
1	0.79	0.98	0.88	193
2	0.87	0.64	0.74	178
3	0.98	0.97	0.97	185
4	0.96	0.92	0.94	209
accuracy			0.90	984
macro avg	0.90	0.90	0.89	984
weighted avg	0.91	0.90	0.90	984

FIGURE 18: RESIDUAL Y-NET CLASSIFICATION REPORT ON DATASET 1

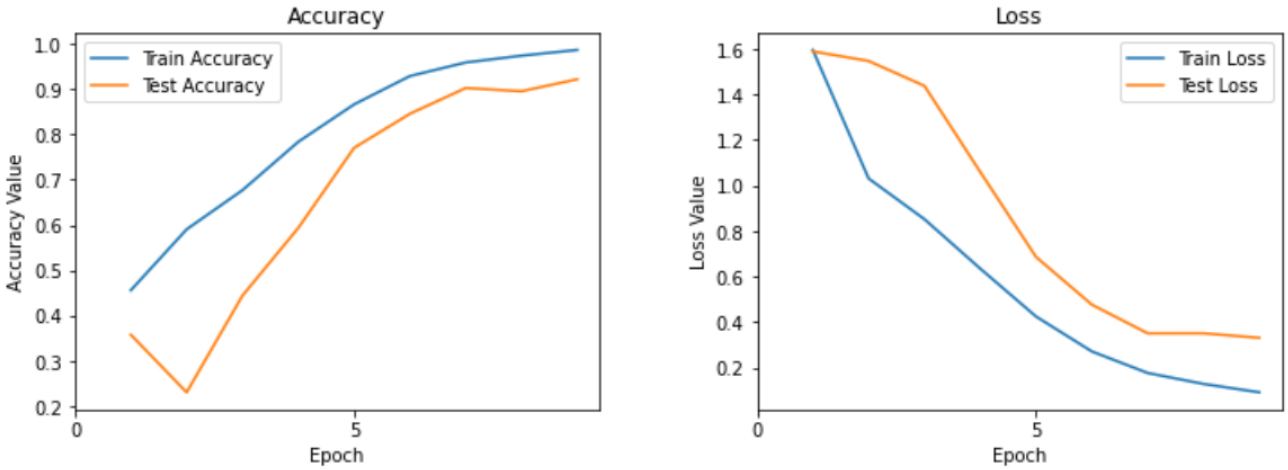


FIGURE 19: RESIDUAL Y-NET ACCURACY AND LOSS ON DATASET 2

Classification Report

	precision	recall	f1-score	support
0	0.95	0.95	0.95	227
1	0.88	1.00	0.94	194
2	0.94	0.73	0.82	179
3	0.93	1.00	0.96	189
4	0.98	0.99	0.98	195
accuracy			0.94	984
macro avg	0.94	0.93	0.93	984
weighted avg	0.94	0.94	0.93	984

FIGURE 20: RESIDUAL Y-NET CLASSIFICATION REPORT ON DATASET 2

The number of epochs is set according to the need of the network. According to the dataset the number of filters and convolutional blocks need to be decided. More number of blocks will result into saturation on the image condensation, since the size of the image is small. According to that, dropouts are put after every layer. Dropouts are added to reduce over-fitting in the model. It ignores some of the visible and non-visible layers, so that the dependency caused by the fully connected layers can be curbed. So basically, the independent learning of individual neurons decreases. After adding required number of dropouts, performance gets better. The training time or the number of epochs also decreases. On Dataset 1 it records train and test accuracy of 96.07% and 86.78% respectively, in 70 epochs (Figure 21). While on Dataset 2, in 25 epochs, the training accuracy reaches to 99.59%. The Test accuracy comes out to be 89.69%. The graph in Figure 23 shows the accuracy for U-net.

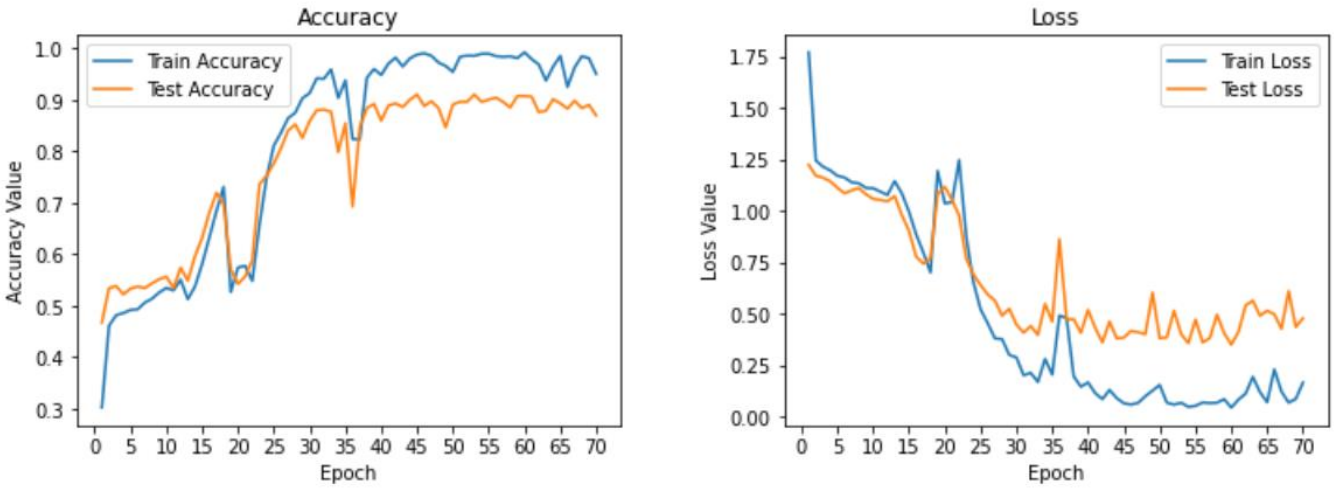


FIGURE 21: U-NET ACCURACY AND LOSS ON DATASET 1

Classification Report

	precision	recall	f1-score	support
0	0.93	0.88	0.91	233
1	0.88	0.92	0.90	201
2	0.82	0.60	0.70	169
3	0.90	0.97	0.93	186
4	0.81	0.94	0.87	195
accuracy			0.87	984
macro avg	0.87	0.86	0.86	984
weighted avg	0.87	0.87	0.87	984

FIGURE 22: U-NET CLASSIFICATION REPORT ON DATASET 1

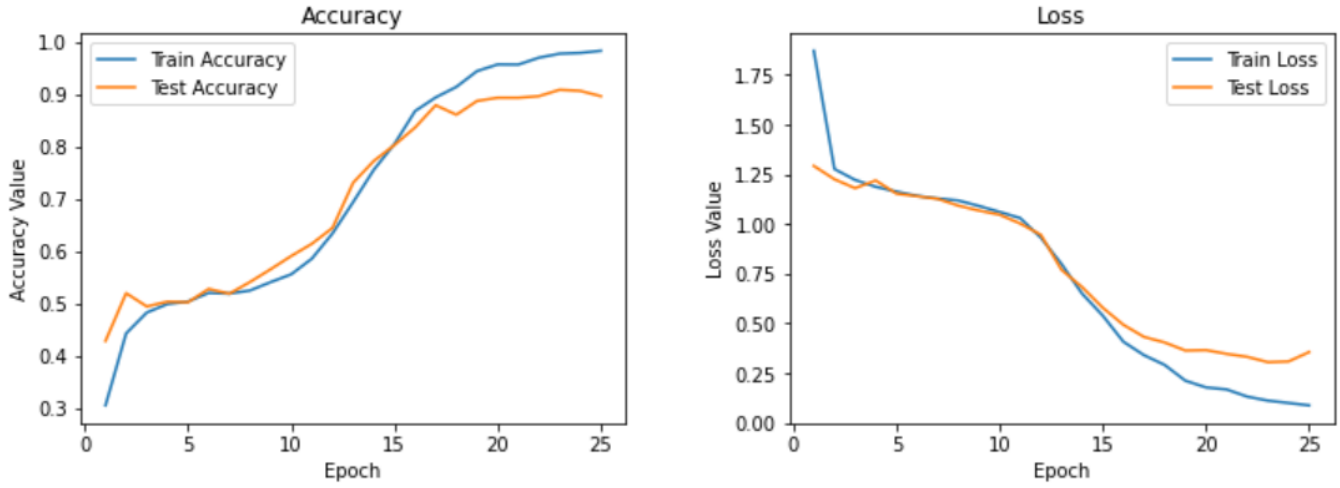


FIGURE 23: U-NET ACCURACY AND LOSS ON DATASET 2

Classification Report

	precision	recall	f1-score	support
0	0.96	0.93	0.95	196
1	0.83	0.98	0.90	196
2	0.94	0.59	0.73	195
3	0.97	0.99	0.98	182
4	0.85	0.99	0.92	215
accuracy			0.90	984
macro avg	0.91	0.90	0.89	984
weighted avg	0.91	0.90	0.89	984

FIGURE 24: U-NET CLASSIFICATION REPORT ON DATASET 2

The setup is kept same for Y-net. Only the number of encoders is two. At the end of residual block, the parallel residual blocks from other encoder is added for feature map collection. The number of epochs was kept high to watch the convergence in training accuracy. Dataset 1 receives 99.82% of train accuracy and 88.75% of test accuracy. Similarly for Y-net, within 25

epochs the training accuracy reaches up to 99.06% but the test accuracy goes up to 90.57%. Since, number of features increased in Y-net, it gives slightly better performance in the neural network.

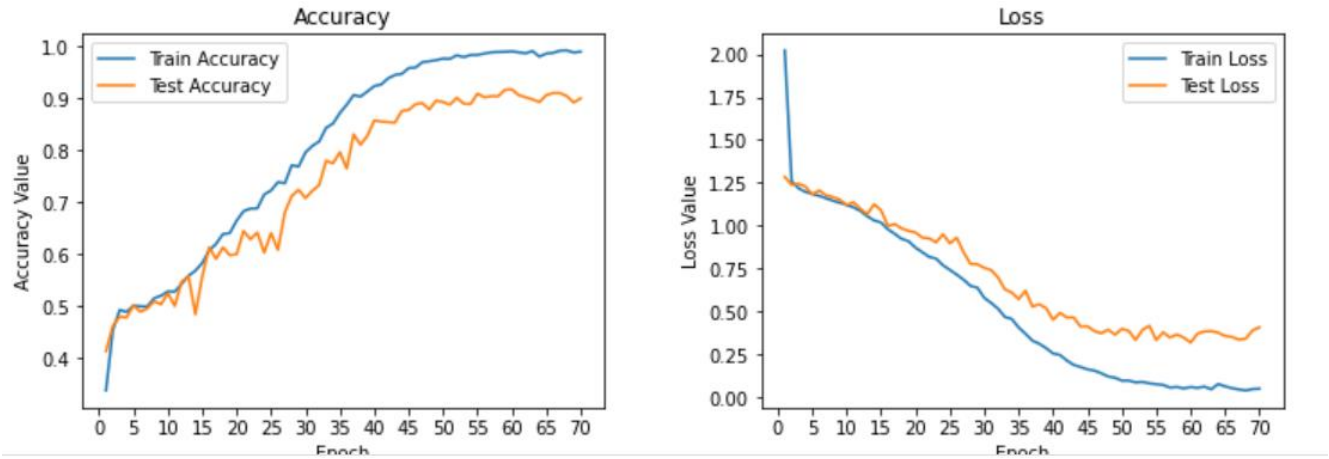


FIGURE 25: Y-NET ACCURACY AND LOSS ON DATASET 1

Classification Report

	precision	recall	f1-score	support
0	0.94	0.94	0.94	224
1	0.83	0.98	0.90	170
2	0.86	0.57	0.69	192
3	0.96	0.98	0.97	206
4	0.84	0.96	0.90	192
accuracy			0.89	984
macro avg	0.89	0.89	0.88	984
weighted avg	0.89	0.89	0.88	984

FIGURE 26: Y-NET CLASSIFICATION REPORT ON DATASET 1

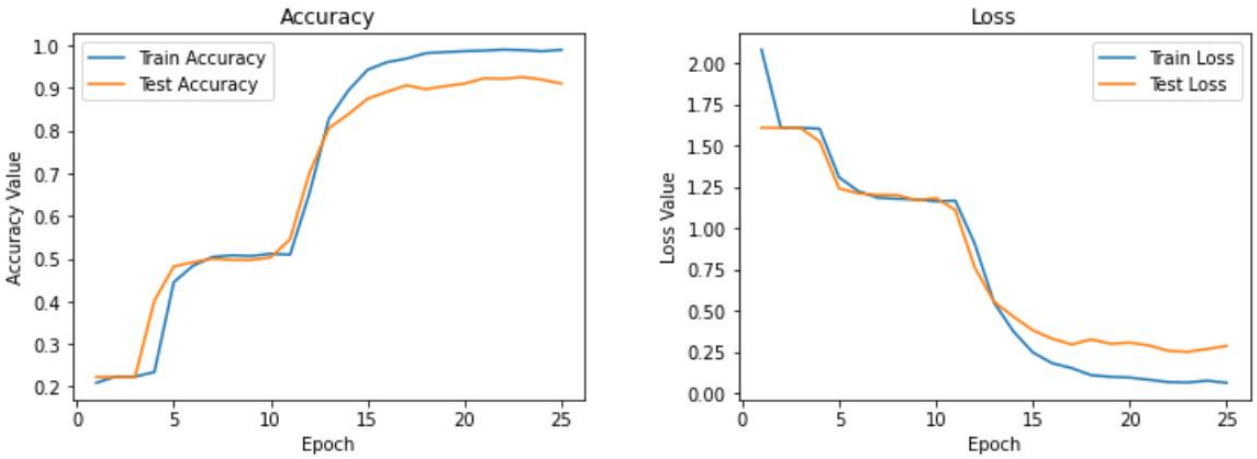


FIGURE 27: Y-NET ACCURACY AND LOSS ON DATASET 2

Classification Report

	precision	recall	f1-score	support
0	0.95	0.93	0.94	219
1	0.87	0.99	0.93	215
2	0.92	0.62	0.74	161
3	0.95	0.97	0.96	196
4	0.87	0.98	0.92	193
accuracy			0.91	984
macro avg	0.91	0.90	0.90	984
weighted avg	0.91	0.91	0.91	984

FIGURE 28: Y-NET CLASSIFICATION REPORT ON DATASET 2

For implementing Deep Residual U-net, the convolutional blocks are replaced with residual blocks. But after adding residual units, a subsequent amount of increase in accuracy is seen. Over Dataset 1, training and test accuracy reaches upto 98.67% and 89.10% respectively (Figure 29). On Dataset 2, in just 15 epochs in Deep residual U-net the training accuracy is

98.89%. The recorded test accuracy is 91.29% (Figure 31).

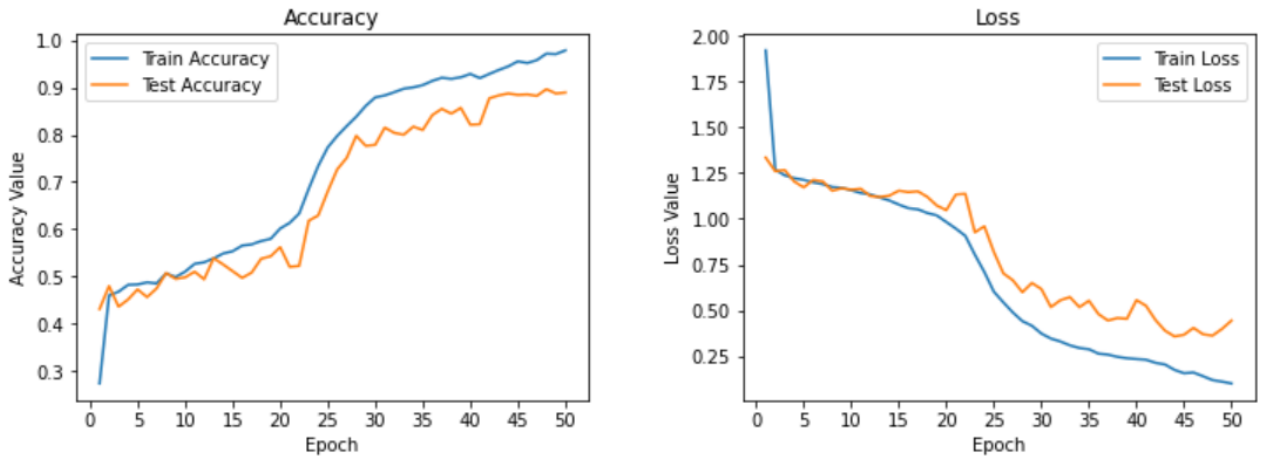


FIGURE 29: RESIDUAL U-NET ACCURACY ON LOSS ON DATASET 1

Classification Report

	precision	recall	f1-score	support
0	0.95	0.90	0.92	232
1	0.84	1.00	0.91	201
2	0.84	0.62	0.72	182
3	0.95	0.98	0.96	192
4	0.88	0.97	0.92	177
accuracy			0.90	984
macro avg	0.89	0.89	0.89	984
weighted avg	0.90	0.90	0.89	984

FIGURE 30: RESIDUAL U-NET CLASSIFICATION REPORT ON DATASET 1

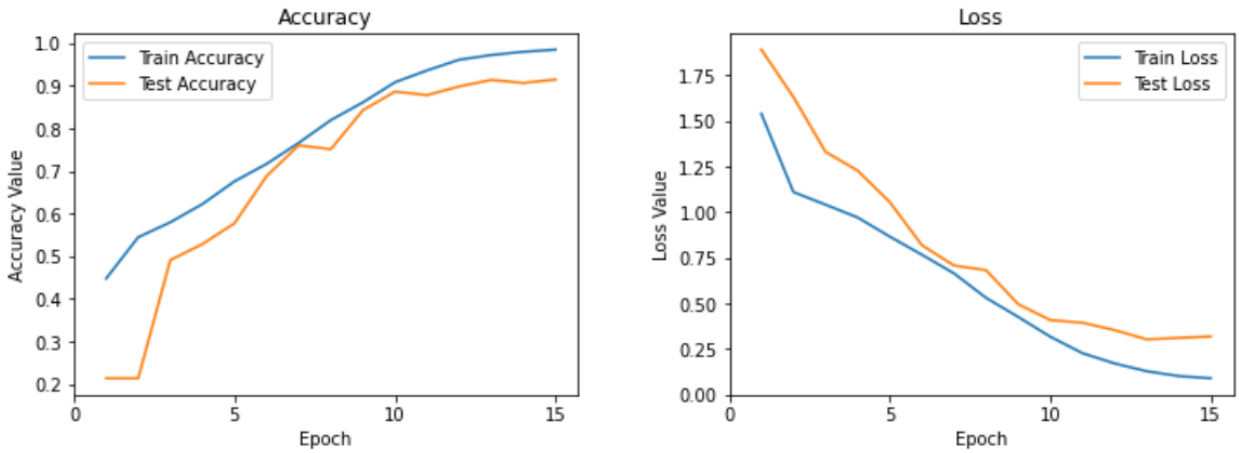


FIGURE 31: RESIDUAL U-NET ACCURACY AND LOSS ON DATASET 2

Classification Report

	precision	recall	f1-score	support
0	0.93	0.95	0.94	225
1	0.81	0.98	0.89	206
2	0.94	0.64	0.76	181
3	0.96	0.99	0.97	184
4	0.96	0.98	0.97	188
accuracy			0.91	984
macro avg	0.92	0.91	0.91	984
weighted avg	0.92	0.91	0.91	984

FIGURE 32: RESIDUAL U-NET CLASSIFICATION REPORT ON DATASET 2

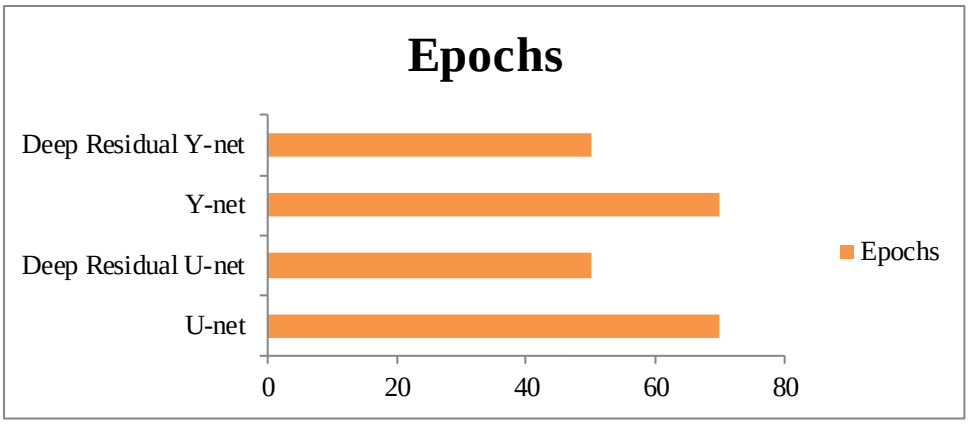


FIGURE 33: EPOCH COMPARISON OF FOUR MODELS ON DATASET 1

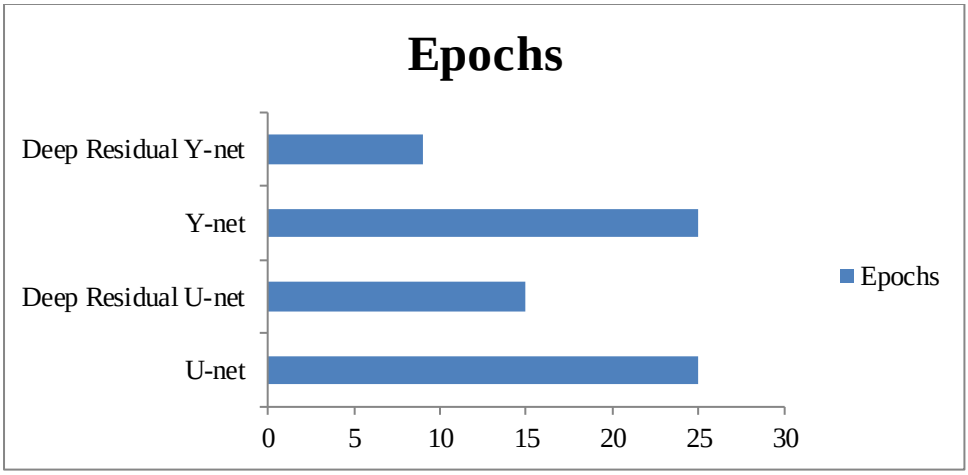


FIGURE 34: EPOCH COMPARISON OF FOUR MODELS ON DATASET 2

The comparison table of performance of different methods is shown in Table 1.

TABLE 1: TEST ACCURACIES

Model	**Dataset 1**	**Dataset 2**
U-net	0.8678	0.8960
Residual U-net	0.8910	0.9129
Y-net	0.8875	0.9057
Residual Y-net	**0.9039**	**0.9360**

The number of epochs for training the model from U-net, Y-net, Residual U-net to Residual Y-net decreases significantly, while the test accuracy increases. Residual Units are proven better instead of traditional convolutional blocks in the networks. Two encoders instead of one encoder help in better feature mapping, and so it produces better segmentation and detection results. The model's performance got better due to the balanced medium size dataset.

CHAPTER 6
CONCLUSION

It is seen that detection and classification accuracy is highly affected by the type of dataset and the working of the neural network model. Results show that balanced dataset gives reliable accuracy than unbalanced because of balanced training of the four different neural network models. Balanced dataset helps to gain comparable amount of feature intelligence on every label or category of the dataset.

It has been noticed that segmentation helps in pixel wise classification which is way better than patch wise classification of images. To compare the performance of the proposed model which inspired from Y-net three other segmentation models are tested upon a balanced and an unbalanced dataset. The proposed method's sole aim is to maintain more features in a condensed form such that segmentation results for image classification can be enhanced when decoding in the decoder. Residual Y-net attained test accuracy of 93.60% in 9 epochs on a medium-size Kaggle Dataset. It does not let to degrade features between low and high levels. Because it uses residual units, the weights of input layers are preserved, and two encoders assisted in the storage of extra features components. There are no restrictions for spatial values. The architecture is flexible in placing any spatial size of residual units. It has been compared with U-net, Y-net and Residual U-net. It outperforms all the three methods.

Future work will include testing the model on a much larger dataset. There will be experiments on bringing more modifications in the architecture to increase the accuracy of classifications. Further research will include ways to tackle unbalanced dataset and give better accuracies.

REFERENCES

[1] W. Guo, M. Li, Y. Dong, H. Zhou, Z. Zhang, C. Tian, R. Qin, H. Wang, Y. Shen, K. Du and others, "Diabetes is a risk factor for the progression and prognosis of COVID-19," *Diabetes/metabolism research and reviews,* vol. 36, p. e3319, 2020.

[2] L. Roncon, M. Zuin, G. Rigatelli and G. Zuliani, "Diabetic patients with COVID-19 infection are at higher risk of ICU admission and poor short-term outcome," *Journal of Clinical Virology,* vol. 127, p. 104354, 2020.

[3] Z. Zheng, F. Peng, B. Xu, J. Zhao, H. Liu, J. Peng, Q. Li, C. Jiang, Y. Zhou, S. Liu and others, "Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis," *Journal of Infection,* 2020.

[4] S. Edagawa, F. Kobayashi, F. Kodama, M. Takada, Y. Itagaki, A. Kodate, K. Bando, K. Sakurai, A. Endo, H. Sageshima and others, "Epidemiological features after emergency declaration in Hokkaido and report of 15 cases of COVID-19 including 3 cases requiring mechanical ventilation," *Global Health & Medicine,* 2020.

[5] L. Orioli, M. P. Hermans, J.-P. Thissen, D. Maiter, B. Vandeleene and J.-C. Yombi, "COVID-19 in diabetic patients: Related risks and specifics of management," in *Annales d'endocrinologie,* 2020.

[6] M. Sen, S. Lahane, T. P. Lahane, R. Parekh and S. G. Honavar, "Mucor in a viral land: a tale of two pathogens," *Indian journal of ophthalmology,* vol. 69, p. 244, 2021.

[7] S. Waizel-Haiat, J. A. Guerrero-Paz, L. Sanchez-Hurtado, S. Calleja-Alarcon and L. Romero-Gutierrez, "A case of fatal rhino-orbital mucormycosis associated with new onset diabetic ketoacidosis and COVID-19," *Cureus,* vol. 13, 2021.

[8] M. Sen, S. G. Honavar, N. Sharma and M. S. Sachdev, "COVID-19 and Eye: A Review of Ophthalmic Manifestations of COVID-19," *Indian Journal of Ophthalmology,* vol. 69, p. 488, 2021.

[9] Y. Sun, R. Zhao, Z. Hu, W. Wang, S. Wang, L. Gao, J. Fei, X. Jian, Y. Li, H. Zheng and others, "Differences in the Clinical and Hematological Characteristics of COVID-19 Patients with and without Type 2 Diabetes," *Journal of diabetes research,* vol. 2020, 2020.

[10] P. K. Reddy, M. S. Kuchay, Y. Mehta and S. K. Mishra, "Diabetic ketoacidosis precipitated by COVID-19: a report of two cases and review of literature," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews,* vol. 14, p. 1459–1462, 2020.

[11] N. E. Palermo, A. R. Sadhu and M. E. McDonnell, "Diabetic ketoacidosis in COVID-19: unique concerns and considerations," *The Journal of Clinical Endocrinology & Metabolism,* vol. 105, p. 2819–2829, 2020.

[12] R. Pal, M. Banerjee, U. Yadav and S. Bhattacharjee, "Clinical profile and outcomes in COVID-19 patients with diabetic ketoacidosis: a systematic review of literature," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews,* vol. 14, p. 1563–1569, 2020.

[13] Q. Shi, X. Zhang, F. Jiang, X. Zhang, N. Hu, C. Bimu, J. Feng, S. Yan, Y. Guan, D. Xu, G. He, C. Chen, X. Xiong, L. Liu, H. Li, J. Tao, Z. Peng and W. Wang, "Clinical Characteristics and Risk Factors for Mortality of COVID-19 Patients With Diabetes in Wuhan, China: A Two-Center, Retrospective Study," *Diabetes Care,* vol. 43, p. 1382–1391, 2020.

[14] A. Odriozola, L. Ortega, L. Martinez, S. Odriozola, A. Torrens, D. Corroleu, S. Martínez, M. Ponce, Y. Meije, M. Presas and others, "Widespread sensory neuropathy in diabetic patients hospitalized with severe COVID-19 infection," *Diabetes research and clinical practice,* vol. 172, p. 108631, 2021.

[15] W. Kerner and J. Brückel, "Definition, classification and diagnosis of diabetes mellitus," *Experimental and clinical endocrinology & diabetes,* vol. 122, p. 384–386, 2014.

[16] A. Misra, "Majorly Resurgent and Uncontrolled Diabetes During COVID19 Era in India Can Be Contained," *Diabetes & Metabolic Syndrome,* 2021.

[17] R. N. Frank, "Diabetic retinopathy," *Progress in Retinal and Eye Research,* vol. 14, pp. 361-392, 1995.

[18] R. Klein, B. E. K. Klein and S. E. Moss, "Visual Impairment in Diabetes," *Ophthalmology,* vol. 91, pp. 1-9, 1984.

[19] S. B. Smith, "Diabetic Retinopathy and the NMDA Receptor.," *Drug news & perspectives,* vol. 15, no. 4, pp. 226-232, 5 2002.

[20] *A Textbook of Clinical Ophthalmology, Third Edition.*, vol. 16, 2003, p. 47.

[21] R. Acharya, C. K. Chua, E. Y. K. Ng, W. Yu and C. Chee, "Application of higher order spectra for the identification of diabetes retinopathy stages," *Journal of medical systems,* vol. 32, p. 481–488, 2008.

[22] F. F. Wahid and G. Raju, "A dual step strategy for retinal thin vessel enhancement/extraction," in *2019 Amity International Conference on Artificial Intelligence (AICAI)*, 2019.

[23] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang and T. Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification," in *2017 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI)*, 2017.

[24] J. J. Gómez-Valverde, A. Antón, G. Fatti, B. Liefers, A. Herranz, A. Santos, C. I. Sánchez and M. J. Ledesma-Carbayo, "Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning," *Biomedical optics express,* vol. 10, p. 892–913, 2019.

[25] A. Samanta, A. Saha, S. C. Satapathy, S. L. Fernandes and Y.-D. Zhang, "Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset," *Pattern Recognition Letters*, vol. 135, p. 293–298, 2020.

[26] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, 2015.

[27] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[28] S. Mehta, E. Mercan, J. Bartlett, D. L. Weaver, J. G. Elmore and L. G. Shapiro, "Y-Net: Joint Segmentation and Classification for Diagnosis of Breast Biopsy Images," *CoRR*, vol. abs/1806.01313, 2018.

[29] Y. Chen, X. Jin, B. Kang, J. Feng and S. Yan, "Sharing Residual Units Through Collective Tensor Factorization in Deep Neural Networks," *CoRR*, vol. abs/1703.02180, 2017.

[30] W. Chen, B. Yang, J. Li and J. Wang, "An Approach to Detecting Diabetic Retinopathy Based on Integrated Shallow Convolutional Neural Networks," *IEEE Access*, vol. 8, p. 178552–178562, 2020.

[31] G. Wang, J. Sun, J. Ma, K. Xu and J. Gu, "Sentiment classification: The contribution of ensemble learning," *Decision Support Systems*, vol. 57, pp. 77-93, 2014.

[32] B. Tymchenko, P. Marchenko and D. Spodarets, "Deep Learning Approach to Diabetic Retinopathy Detection," *CoRR*, vol. abs/2003.02261, 2020.

[33] R. Patel and A. Chaware, "Transfer Learning with Fine-Tuned MobileNetV2 for Diabetic Retinopathy," in *2020 International Conference for Emerging Technology (INCET)*, 2020.

[34] Z. Zeng, W. Xie, Y. Zhang and Y. Lu, "RIC-Unet: An improved neural network based on Unet for nuclei segmentation in histology images," *Ieee Access*, vol. 7, p. 21420–21428, 2019.

[35] X. Xiao, S. Lian, Z. Luo and S. Li, "Weighted res-unet for high-quality retina vessel segmentation," in *2018 9th international conference on information technology in medicine and education (ITME)*, 2018.

[36] J. Hu, H. Wang, S. Gao, M. Bao, T. Liu, Y. Wang and J. Zhang, "S-unet: A bridge-style u-net framework with a saliency mechanism for retinal vessel segmentation," *IEEE Access*, vol. 7, p. 174167–174177, 2019.

[37] S. Suri, A. Gupta and K. Sharma, "Comparative Analysis of Ranking Algorithms Used On Web," *Annals of Emerging Technologies in Computing (AETiC)*, vol. 4, 2020.

[38] A. K. Tripathi, K. Sharma, M. Bala, A. Kumar, V. G. Menon and A. K. Bashir, "A Parallel Military-Dog-Based Algorithm for Clustering Big Data in Cognitive Industrial Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 17, pp. 2134-2142, 2021.

[39] S. Woo, J. Park, J.-Y. Lee and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018.

[40] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen and C. Fan, "Sa-unet: Spatial attention u-net for retinal vessel segmentation," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021.

[41] Y. Zong, J. Chen, L. Yang, S. Tao, C. Aoma, J. Zhao and S. Wang, "U-net Based Method for Automatic Hard Exudates Segmentation in Fundus Images Using Inception Module and Residual Connection," *IEEE Access*, vol. 8, p. 167225–167235, 2020.

[42] C. Guo, M. Szemenyei, Y. Hu, W. Wang, W. Zhou and Y. Yi, "Channel Attention Residual U-Net for Retinal Vessel Segmentation," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021.

[43] A. Foo, W. Hsu, M. L. Lee, G. Lim and T. Y. Wong, "Multi-Task Learning for Diabetic Retinopathy Grading and Lesion Segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

[44] J. E. McManigle, R. R. Bartz and L. Carin, "Y-Net for Chest X-Ray Preprocessing: Simultaneous Classification of Geometry and Segmentation of Annotations," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020.

[45] A. Mohammed, S. Yildirim, I. Farup, M. Pedersen and Ø. Hovde, "Y-net: A deep convolutional neural network for polyp detection," *arXiv preprint arXiv:1806.01907*, 2018.

[46] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang and T. Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification," in *2017 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI)*, 2017.

[47] S. Masood, T. Luthra, H. Sundriyal and M. Ahmed, "Identification of diabetic retinopathy in eye images using transfer learning," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, 2017.

[48] J. Dash and N. Bhoi, "An unsupervised approach for extraction of blood vessels from fundus images," *Journal of digital imaging*, vol. 31, p. 857–868, 2018.

[49] N. Singh, L. Kaur and K. Singh, "Histogram equalization techniques for enhancement of low radiance retinal images for early detection of diabetic retinopathy," *Engineering Science and Technology, an International Journal*, vol. 22, p. 736–745, 2019.

[50] Z. Jiang, H. Zhang, Y. Wang and S.-B. Ko, "Retinal blood vessel segmentation using fully convolutional network with transfer learning," *Computerized Medical Imaging and Graphics*, vol. 68, p. 1–15, 2018.

[51] S. H. Bhat and P. Kumar, "Segmentation of optic disc by localized active contour model in

retinal fundus image," in *Smart Innovations in Communication and Computational Sciences*, Springer, 2019, p. 35–44.

[52] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros and others, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, p. 2402–2410, 2016.

[53] K. M. Adal, P. G. Van Etten, J. P. Martinez, K. W. Rouwen, K. A. Vermeer and L. J. van Vliet, "An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus images," *IEEE transactions on biomedical engineering*, vol. 65, p. 1382–1390, 2017.

[54] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *nature*, vol. 521, p. 436–444, 2015.

[55] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE transactions on medical imaging*, vol. 35, p. 1299–1312, 2016.

[56] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami and M. K. Khan, "Medical image analysis using convolutional neural networks: a review," *Journal of medical systems*, vol. 42, p. 1–13, 2018.

[57] Z. Zhang, Q. Liu and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, p. 749–753, 2018.

[58] D. Han, J. Kim and J. Kim, "Deep pyramidal residual networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.

[59] K. Cao and X. Zhang, "An improved res-unet model for tree species classification using airborne high-resolution images," *Remote Sensing*, vol. 12, p. 1128, 2020.

[60] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.

[61] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, p. 1–48, 2019.

LIST OF PUBLICATIONS OF THE CANDIDATE'S WORK

[1] K. S. Anindita Roy, "Residual Y-net for Detection of Diabetic Retinopathy," in *Innoative Computing,Intelligent Communication and Smart Electrical Systems*, 2021. [Accepted]