CLASSIFICATION OF BREAST CANCER WITH ARTIFICIAL NEURO-FUZZY INFERENCE SYSTEM (ANFIS)

MARGARET OMAMUROMU ASHEDAGHO

PSC1908815

FALCULTY OF PHYSICAL SCIENCE

UNIVERSITY OF BENIN

BENIN CITY, EDO STATE.

MARCH, 2024

CLASSIFICATION OF BREAST CANCER WITH ARTIFICIAL NEURO-FUZZY INFERENCE SYSTEM (ANFIS)

MARGARET OMAMUROMU ASHEDAGHO

PSC1908815

BEING A PROJECT PROPOSAL SUBMITTED TO THE FALCULTY OF PHYSICAL SCIENCES, IN PARTIAL FULFILMENT OF THE REQUIREMNET OF COMPUTER SCIENCE OF NIGERIA FOR THE AWARD OF BACHELOR OF SCIENCE UNIVERSITY OF BENIN, BENIN CITY, EDO STATE, NIGERIA.

MARCH, 2024

APPROVAL

This project report written by **MARGARET OMAMUROMU ASHEDAGHO** with matriculation number **PSC1908815,** in partial fulfillment of the requirement for the award of the University of Benin Bachelor of Science (B.Sc.) degree in Computer Science, is adequate both in scope and content and it is hereby approved for presentation.

_____ _____

PROF. G. O. EKUOBASE, PhD **Date**

(Head of Department)

DECLARATION

This is to declare that this research project titled **CLASSIFICATION OF BREAST CANCER WITH ARTIFICIAL NEURO-FUZZY INFERENCE SYSTEM (ANFIS)** is solely the result of my work except where acknowledge are being derived from other resources or person.

_____ _____

MARGARET OMAMUROMU ASHEDAGHO **DATE**

PSC1908815

CERTIFICATION

This is to certify that this research project titled **CLASSIFICATION OF BREAST CANCER WITH ARTIFICIAL NEURO-FUZZY INFERENCE SYSTEM (ANFIS)** was carried at by Margaret Omamuromu Ashedagho with examination number PSC1908815 in the DEPARTMENT OF COMPUTER SCIENCE under the supervision of Dr. E. C. Igodan has been prepared in accordance with the regulation governing the preparation of project in the Faculty of physical sciences, University of Benin, Benin City, Edo State.

Signature: _____ _____

Dr. E. C. Igodan **Date**

(Project Supervisor)

Signature: _____ _____

PROF. G. O. EKUOBASE, PhD **Date**

(Head of Department)

Signature: _____ _____

Name: _____ **Date**

(Chief Examiner)

DEDICATION

This work is dedicated to my family, both home and abroad for their support.

ACKNOWLEDGEMENT

My utmost acknowledgement goes to God Almighty, for seeing me through all my years in school and for completing this good work that He started in my life. I would like to express my profound gratitude to my project supervisor, Dr E. C. Igodan for his consistent guidance throughout the period of this project. I will also like to specially thank my project coordinator, and all the other lecturers in the Department of Computer Science, University of Benin, for their support and inspiration in one way or the other; especially Prof. Godspower O. Ekuobase (services computing), You have all worked hard to set me on the proper road in my professional pursuit and to implant in me sound knowledge about important parts of life.

I want to personally appreciate as many who contributed to the success of my undergraduate studies and the completion of this project, be it in moral support, financial assistance or spiritual mentoring. Pastor praise, Pastor (Dr) Erivona Obaro.

Finally, I would also like to thank my father Mr Godwin O. Ashedagho, my siblings especially Miss Sarah O. Ashedagho and Miss (Barr) Elizabeth E. Ashedagho. Thank you so much for the love, encouragement, financial support and moral support. My friends most especially, Miss Joy Adamma, Miss Marcellina O. Ugege for your continuous prayers and words of encouragement throughout the project. I'm so grateful.

ABSTRACT

Breast cancer stays one of the maximum standard and life-threatening illnesses affecting girls globally. Accurate and early diagnosis is critical for effective treatment and improved survival rates. This project explores the application of an Artificial Neuro-Fuzzy Inference System (ANFIS) for the classification of breast cancer. ANFIS combines the learning capabilities of neural networks with the reasoning capabilities of fuzzy logic, creating a hybrid model that can handle the complexities and uncertainties inherent in medical data.

The research involves the collection and preprocessing of breast cancer datasets, followed by the design and implementation of an ANFIS model. The model is trained using a portion of the dataset and tested on the remaining data to evaluate its classification performance. Key performance metrics such as accuracy, sensitivity, specificity, and the area under the receiver operating characteristic (ROC) curve are used to assess the effectiveness of the ANFIS model.

Preliminary results indicate that the ANFIS model demonstrates promising accuracy in distinguishing between benign and malignant breast tumors. The adaptive learning process of the ANFIS allows for continuous improvement and adjustment of the model, enhancing its diagnostic capabilities over time. This study highlights the potential of ANFIS as a reliable and efficient tool for breast cancer classification, contributing to the advancement of artificial intelligence applications in medical diagnostics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND STUDY

Breast cancer is the most commonly diagnosed cancer and the leading cause of death among female patients, which seriously threatens the health of women in the whole world (Wang and Wu, 2023). The breasts of a woman are crafted from fat, supportive (connective) tissue and tissues with glands called lobes. These lobes are milk glands in which breast milk is produced. These are related to the nipple through a community of milk ducts. The frame is made from billions of tiny cells. Normally, cells develop and multiply in a tightly regulated fashion. New cells are only made when and where they are needed. When most cancers happens the cells increase cycle is going haywire making them multiply uncontrollably. This leads to formation of a lump that may be benign or non-cancerous or aggressive if its growth – termed malignant or cancerous (Mandal, 2023).

Cancer is a broad term for a class of diseases characterized by abnormal cells that grow and invade healthy cells in the body. Cancer starts offevolved withinside the cells that are the simple constructing blocks that make up tissue. Tissue is determined withinside the breast and different components of the body. Sometimes, the technique of mobileular boom is going incorrect and new cells shape whilst the frame doesn't want them and antique or broken cells do now no longer die as they should. When this occurs, a increase of cells frequently paperwork a mass of tissue referred to as a lump, growth, or tumor. (Shockney, 2023).

Breast most cancers happens while malignant tumors broaden withinside the breast. These cells can unfold with the aid of using breaking farfar from the unique tumor and coming into blood vessels or lymph vessels, which department into tissues in the course of the body.

When most cancers cells journey to different elements of the frame and start adverse different tissues and organs, the method is referred to as metastasis. (Shockney, 2023).

Breast most cancers is a especially heterogeneous sickness this is advanced with the aid of using mutual effect of genetic chance elements and environmental elements. It results in modern aggregation of genetic and epigenetic adjustments in breast most cancers cells. Although epidemiological evidence highlights the presence of risk factors (such as age, obesity, alcohol use, and exposure to estrogen in lifetime), family history of breast cancer is the strongest one (Attaholahi et al., 2015).

Differentiating a malignant tumor from a benign one is a totally tedious project because of the structural similarities among the two.

It is an exceptionally essential and time eating mission for the doctor to correctly discover the structural differences. Computer aided analysis structures are essential for sample recognition , aiming to help docs in making analysis decisions. Machine learning has been successfully applied in computer-aided diagnosis (CAD) systems. These methods learn hypotheses from a large number of diagnosed samples, i.e the data collected from a number of patients along with the corresponding diagnoses made by medical experts, to assist the medical experts in making a diagnosis in the future (Fatima and Amine, 2012].

Treatment of breast most cancers is complicated and entails a aggregate of various modalities such as surgery, radiotherapy, chemotherapy, hormonal therapy, or organic treatment options introduced in various sequences. (Lukasiewicz et al., 2021).

Artificial intelligence is defined as the ability of a computer to learn algorithms to reason and perform tasks including reading, writing, interacting, problem-solving, and decision-making (Zheng et al., 2022). Algorithms and networks that power artificial intelligence are referred to

as machine learning.ML, a subfield of AI, is mainly used to extract features from training set data and build Analytical mathematical fashions for prediction and evaluation of unknown data. It can help you identify patterns from experiences, environment and trials (Zheng, 2022). Some of the commonly used machine learning algorithms are Artificial neuro fuzzy inference system which comprises fuzzy inference system and neural networks. Fuzzy inference system is a complex system model with learning capabilities. It is a type of artificial intelligence that uses fuzzy logic to make sense of ambiguous or inaccurate data. Fuzzy logic is a mathematical framework for reasoning with vague or ambiguous information by employing degrees of accuracy rather than binary true/false results. Artificial neural networks are a type of machine learning algorithm that is inspired by the structure and operation of the human brain. They consist of interconnected nodes that analyze and transmit data. They are capable of learning complex patterns and connections in large datasets, it makes them ideal for classifying medical diagnosis (Canayaz, 2019).

ANN has studying capabilities, however it can not interpret results. That is, it acts as a black boxes. Whereas, Fuzzy Inference System (FIS) can interpret the results using a rule base, it cannot learn. To conquer the dangers of each techniques, joint use of ANN and FIS may be hired to get higher results (Shah, 2023).

Adaptive Neuro-Fuzzy Inference System (ANFIS) were efficiently implemented to rule-primarily based totally system controls, class tasks, feature approximation, and sample reputation problems. (Sada and Ikpeseni, 2021).

1.2 RESEARCH MOTIVATION

The diagnosing of breast cancer is a complex task that necessitates analyzing large volumes of medical data including mammogram images and patients records. There are limitations to traditional diagnostic methods of breast cancer which can be rectified with computer aided

diagnostic methods. Traditional diagnostic approaches, including mammography and biopsy, have proven useful, but they are invasive, time-consuming, and costly. Not every cancer can be detected by mammography. Some may appear between two mammograms or are not evident on the scan. There lies the risk of overdiagnosis, since it's difficult to differentiate the malignancy of a cancer at the early stages of screening. The overwhelming number of images, which lead to a heavy workload for radiologists and a sluggish reporting period, suggests the need for computer-aided detection techniques and platforms (Zheng et al., 2022).

1.3 AIM AND OBJECTIVES OF STUDY

The aim of this project is to design a predictive model for the classification of breast cancer using ANFIS.

The specific objectives are:

(a) Carefully use a feature selector to select its distinctive features.

(b) Design an ANFIS model using (a) above.

(c) Implement (b) using Python programming language.

(d) Carry out a case study of breast cancer.

1.4 RESEARCH METHODOLOGY

The research methodology includes the following steps:

1. **Data Collection:** It involves gathering data on the clinical, pathological and molecular aspects of breast cancer tumors which would be used as input features in the ANFIS model.

2. **Data Preprocessing:** The acquired data will be preprocessed, which is called data cleaning, to eliminate any missing values or other abnormalities that could affect the model's performance.The attribute missing values are replaced by its mean value.

3. **Data Splitting:** The preprocessed data will be split into training and testing sets.The training set will be used to train the ANFIS model while the testing set will be used to assess and evaluate its performance.

4. **ANFIS Model Development:** ANFIS will be used to model the relationship between the input features and class labels (breast cancer subtypes).

5. **Model Evaluation:** The model's performance will be evaluated using a variety of performance indicators including accuracy, precision, recall and the F1 score.

6. The ANFIS model's performance will be compared with other breast cancer classification models to determine its reliability.

7. Finally, assess and draw conclusions regarding the ANFIS model's performance.

1.5 SIGNIFICANCE OF STUDY

The benefits of this study shows that:

1. This proposed classifier outperforms some other machine learning algorithms and current methods in terms of dataset classification.

2. It also makes a significant contribution to the performance improvement of traditional classification algorithms in data mining research.

3. The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a computational model that combines neural network adaptability and fuzzy logic interpretability.

1.6 SCOPE OF STUDY

To appropriately categorize breast cancer subtypes, the model will use a variety of clinical and pathological data. In this study, model, the materials and methods used include:

1. Python programming language is used to develop the ANFIS model.
2. A deep learning architecture; comprising of neural network and fuzzy logic (an ANFIS model)
3. Wisconsin breast cancer dataset.
4. Feature selectors; which are information gain (IG), gain ratio (GR), chi-square (CS); which are used to select the features to train the model.
5. Ensemble method which was majority voting.

CHAPTER TWO

LITERATURE REVIEW

2.0 STRUCTURE OF THE BREAST

The breasts are paired systems positioned at the anterior thoracic wall, withinside the pectoral region. They are found in each adult males and ladies, but are extra outstanding in ladies following puberty. (Kammath, 2022). Males and females have slightly distinct breast architecture. The glandular tissue and milk ducts found in female breasts facilitate nursing. (Cleveland clinic). The breasts of an person girl are milk-generating glands at the the front of the chest wall. They relaxation at the pectoralis principal and are supported with the aid of using and connected to the the front of the chest wall on both aspect of the sternum with the aid of using ligaments. Each breast carries 15-20 lobes organized in a round fashion. The fats that covers the lobes offers the breast its length and shape. Each lobe incorporates many lobules, at the surrender of which is probably glands that produce milk in response to hormones (see the image below). (Chalasani, 2023).

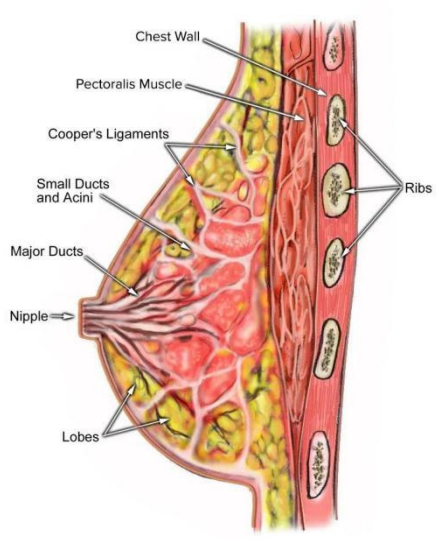

Fig 2.1: STRUCTURE OF THE BREAST

The majority of the breast includes glandular (milk-producing) and fatty tissues. However, the ratio of the glandular to fatty tissue varies amongst individuals. It also contains fatty and connective tissues that provide structure and support. The breast is closely motivated through the intercourse hormone estrogen. As menopause approaches, the tiers of estrogen declines which additionally decreases the glandular tissues. (Rivard et al., 2023).

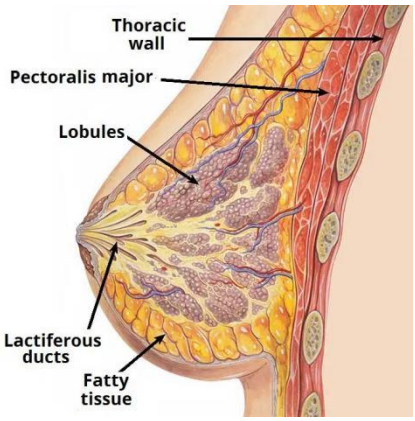

Fig 2.2 the internal structure of the breast. By Patrick J. Lynch, via Wikimedia Commons.

2.1 FUNCTIONS OF THE BREAST

The major function of the breast is to produce milk to nourish infants. It also produces sexual attraction and arousal. The nerve endings, when stimulated during sexual activity, produce pleasurable sensations. Female hormones, including estrogen, progesterone, and prolactin, play critical roles in breast development and function. Estrogen expands milk ducts, allowing them to form side branches that can convey more milk. Prolactin stimulates the production of progesterone and prepares glands for milk production. Progesterone stimulates the quantity and size of lobules in preparation for nursing. Following ovulation, this hormone enlarges blood vessels and breast cells. Oxytocin promotes the release (or ejection) of breast milk. (Cleveland clinic).

2.1.2 OVERVIEW OF BREAST CANCER

Cancer begins offevolved whilst cells start to develop out of control. It is a disease in which body cells develop abnormally. It can originate anywhere and spread to other parts of the body. Human cells normally expand and multiply (by a process known as cell division) to produce new cells as needed by the body. When cells get vintage or damaged, they die and are changed with the aid of using new cells. This ordered process occasionally fails, allowing abnormal or damaged cells to grow and reproduce when and where they should not to form a lump or tumor. Tumors can be malignant or benign. Cancerous tumors infect neighboring tissues and can move to distant parts of the body to generate new tumors (a process known as metastasis). Cancerous tumors also are called malignant tumors. Many cancers develop into solid tumors, while blood cancers, such as leukemias, do not. Benign tumors do now no longer unfold or infect surrounding tissues. Benign tumors seldom recur after removal, although malignant ones occasionally do. However, benign tumors can come to be extraordinarily enormous. (National cancer institute).

Increased occurrence of most cancers in latest years and its effect on one-of-a-kind physical, mental, and social dimensions of human lifestyles have became it into a first-rate hassle of the century. The incidence of this disease in developed countries varies from 1 to 2 percent, with an almost 5% yearly increase in less developed countries. According to estimates, extra than 7 million human beings globally die from cancer.

Breast cancer is a cancer that develops in the breast tissue that primarily affects the inner layer of milk glands or lobules and ducts (tiny tubes that transport milk). It is the second greatest cause of cancer-related death and the main cause of death for women between the ages of 45 and 55. It occurs when cells in the breast grow uncontrolled, resulting in a

tumor that can infiltrate surrounding tissues and potentially spread to other parts of the body. It can begin in a single or each breasts.

2.2 RISK FACTORS/CAUSES OF BREAST CANCER

Anything that raises your probability of contracting a disease, like cancer, is a risk factor. However, having one or greater threat elements does now no longer assure that you'll increase the illness. Certain breast cancer risk factors are unavoidable, such as age and family history, but there are other risk factors that you can manage.

Non-Modifiable risk factors of breast cancer includes:

a. OLD AGE

The aging population's increased risk of cancer can be ascribed to two processes: gradually accruing DNA damage and a steady reduction in host defenses against tumor growth. DNA damage results from cumulative exposure to carcinogenic chemicals, radiation, and viruses. Another type of damage is the cumulative impact of endogenous activities that produce reactive oxygen species, which are extremely harmful to cellular structures. Aging is connected with decreased synthesis and activity of protective enzymes and hormones capable of repairing or bypassing damaged DNA, as well as immune defenses against viruses and tumor growth. (Tesarova, 2012).

b. FAMILY HISTORY

Breast most cancers and ovarian most cancers are lady hormone-responsive cancers which can be carefully associated with one another. Approximately 5-10% of newly diagnosed breast cancer patients have a family history of breast or ovarian cancer, indicating the importance of genetic or non-genetic inheritance in the development of breast cancer. Several well-known genes have been found to be enriched in families with

breast or ovarian cancer. For example, 50-85% of women with mutations in the breast cancer type 1 and 2 genes (BRCA1 and BRCA2) may develop breast and/or ovarian cancer during their lifetime. Mutations in these frequent genes cause breast and ovarian malignancies. Breast/ovarian cancers caused by mutations in these frequent genes are referred to as Hereditary Breast and Ovarian Cancer (HBOC), because mutations can be passed down through family members and cancers are more prevalent in these families (Liu et al., 2021).

c. GENETIC MUTATIONS

BRCA1 (chromosome 17) and BRCA2 (chromosome 13) are two key genes known for their high penetrance. They are mostly associated with the higher risk of breast carcinogenesis.

Carcinogenesis is a complex process that is primarily influenced by both genetic predisposition and environmental factors. It can arise in any cell, tissue, or organ, causing pathological changes that result in a wide range of malignancies. (Łukasiewicz et al., 2021).

d. RACE/ETHNICITY

Breast cancer incidence and mortality rates vary greatly by geographic region, racial and ethnic group. For example, breast cancer incidence rates are greater, whereas mortality rates are lower in transitioned than transitional countries.

e. PREGNANCY AND BREASTFEEDING

Breastfeeding causes hormonal changes in the breast, potentially lowering the chance of developing breast cancer. Both pregnancy and breastfeeding limit the number of menstrual cycles which reduces exposure to certain hormones linked to an elevated risk

of luminal type breast cancer. Nulliparity is a well-established risk factor for luminal type breast cancer; "Nulliparity" is a medical term for a woman who has not given birth to a child.It does not necessary imply that she has never been pregnant; someone who has experienced a miscarriage, stillbirth, or elective abortion but has never given birth to a live child is still considered nulliparous (Nwadike, 2020). Endogenous hormone exposure, particularly estrogen and progesterone, raises the risk of breast cancer in women significantly. The occurrence of events such as pregnancy, nursing, first menstruation, and menopause, as well as their duration and associated hormonal imbalance, are critical in terms of possible carcinogenic events in the breast microenvironment. The first full-term pregnancy at a young age (particularly in the early twenties), followed by an increasing number of births, is connected with a lower risk of breast cancer. The dysregulated hormone levels during preeclampsia, which include increased progesterone and decreased estrogen levels, as well as insulin, cortisol, insulin-like growth factor-1, androgens, human chorionic gonadotropin, corticotropin-releasing factor, and protein deviating from physiological ranges, have a protective effect against breast cancer. Breastfeeding for a longer period reduces the incidence of both ER/PR-positive and -negative malignancies (Łukasiewicz et al., 2021).

f. MENSTRUAL PERIOD AND MENOPAUSE

The biological explanation for this association is the early and prolonged exposure of the breast epithelium to estrogens produced during the period of activity of the ovaries. Pregnancy has a protective effect on breast cancer, such as decreased levels of estrogen and progesterone, increased levels of sex hormone-binding globulin, and pregnancy-induced differentiation of breast tissue. The fact that, at birth, the mammary epithelial cells, which have a high degree of terminal differentiation, can metabolize toxins and repair DNA damage more effectively. Late menopausal age is a well-established risk

factor for developing breast cancer. Women at a late age at menopause may have a higher risk of developing breast cancer due to the longer duration and higher amount of exposure to estrogen and progesterone they experience (Khalis et al., 2018).

g. INSUFFICIENT VITAMINS SUPPLEMENTATION

Vitamins are thought to influence breast cancer risk via a variety of processes, including antioxidant activity against free radical-induced DNA damage, cellular integrity maintenance, coenzyme and immunological system functions, and reproductive system functions. Although, in a case study, Vitamin E supplementation was linked to a lower incidence of breast cancer among women with low dietary intakes. It was previously proposed that the effect of supplemental vitamin E intake on breast cancer risk may be based on the dietary intake level within the population or maybe food sources (Dorjgochoo et al., 2009).

h. DENSITY OF BREAST TISSUE

Collagen, fat, and epithelial and stromal cells are the tissue constituents that lead to differences in PMD (percent mammographic density). Heritable factors have a major role in determining these breast tissue components. Moreover, every constituent possesses characteristics that could impact the likelihood and advancement of breast cancer. The risk of breast cancer is closely linked to PMD, and factors influencing PMD may also contribute to the causes of breast cancer. Determining the factors influencing PMD may also help identify factors that can lower the incidence of breast cancer. Only 20% to 30% of the variation in PMD found in the community can be attributed to age, parity, or menopausal state; the remainder, or the heritability, of PMD may be explained by genetic variables (Boyd et al., 2011). The larger breast tissue density correlates with the increased

breast cancer risk; this pattern is evident both in premenopausal and postmenopausal women (Łukasiewicz et al., 2021).

Modifiable Factors of breast cancer includes:

a. HORMONAL REPLACEMENT THERAPY

Hormone replacement therapy (HRT), also known as estrogen replacement therapy, menopausal hormone therapy, or post-menopausal hormone therapy, can be an effective treatment for menopausal symptoms. However, it is crucial to note that some kinds of HRT (hormone replacement therapy) can raise your risk of breast cancer. These hazards vary based on factors such as the type of HRT (hormone replacement therapy), the dose you take, your age when you start taking it, and the duration of use (DePolo, 2024).

b. DIETHYLSTILBESTROL

Diethylstilbestrol (DES) is a synthetic version of the female hormone estrogen. It was recommended to pregnant women between 1940 and 1971 to avoid miscarriage, early labor, and other pregnancy problems. The use of DES decreased as research in the 1950s revealed that it was ineffective in preventing these disorders, however it was still used to end lactation, for emergency contraception, and to treat menopausal symptoms in women. DES is now recognized as an endocrine disruptive agent, one of several compounds that might interact with the endocrine system, potentially causing cancer, birth defects, and other developmental disorders.

Females exposed to DES in utero, often known as DES daughters, are at an elevated risk of breast cancer.

In a study, it was discovered that postmenopausal DES daughters showed altered estrogen metabolism, implying that prenatal exposure to this endocrine disruptor can influence estrogen metabolism for many years.DES daughters may be more likely to have breast cancer after the age of 40. A 2006 study from the United States revealed that breast cancer risk is not increased in DES daughters overall, but that after age 40, DES daughters have nearly twice the chance of breast cancer as unexposed women of the same age and with comparable risk (National Cancer Institute, 2021).

c. PHYSICAL ACTIVITY

The Women who aren't bodily lively have a better danger of having breast cancer. It can be because of its outcomes on frame weight, inflammation, and hormone levels. Physical activity is most beneficial to women during or after menopause, as well as to those with upper-normal to overweight body compositions (waist circumference, body fat, or BMI). The differences in physical activity levels between women who developed breast cancer and those who did not were most pronounced in perimenopausal and postmenopausal women, women with lower-middle and upper-middle waist circumferences, women with upper-normal and overweight BMI, and lower- and upper-middle body fat percentages (Boraka et al., 2022).

d. OVERWEIGHT/OBESITY

Obesity protects against breast cancer in premenopausal women, whereas post-menopausal women who are obese are more likely to acquire breast cancer (Dehesh et al., 2023).

2.3 COMMON SIGNS AND SYMPTOMS OF BREAST CANCER

The signs and signs of breast most cancers might also additionally include:

1. A breast lump or thickened location of pores and skin that feels extraordinary from encompassing tissue. If a palpable lump is found and possesses any of the following features, breast cancer may be present:

a. Hardness.

b. Irregularity.

c. Focal nodularity.

d. Fixation to skin or muscle.

2. The nipple that appears flattened or turns inward.

3. Changes in the color of the breast skin. In human beings with white pores and skin, the breast pores and skin can also additionally appearance red or red. In human beings with brown and black pores and skin, the breast pores and skin may also appearance darker than the alternative pores and skin at the chest or it could appearance pink or purple.

4. Change in the size, shape or appearance of a breast.

5. Dilated veins.

6. Nipple discharge.

7. Edema or peau d'orange.

8. Changes to the pores and skin over the breast, inclusive of pores and skin that appears dimpled or seems like an orange peel.Peeling, scaling, crusting or flaking of the pores and skin at the breast. (Mayo Clinic).

2.3.1 EARLY DETECTION OF BREAST CANCER

Early detection of breast cancer improves treatment options of breast cancer and increases the chances of survival significantly (Ginsburg et al., 2021). As patients with smaller tumors at diagnosis had a considerably reduced chance of dying and a higher survival rate, early detection of breast cancer is essential for effective treatment and a better prognosis. It is the main line of defense against metastatic breast cancer. There are several methods to detect irregularities or abnormalities in the breast, which includes:

a. **Self-Examination of the Breast (BSE):** It is recommended that women self-examine their breasts on a regular basis to become acquainted with how their breasts should feel and look. Breast self examination (BSE) includes examining the skin for any changes, thickening, or lumps, and reporting any anomalies as soon as possible to a healthcare professional (Thomas et al., 2002).

b. Clinical Breast Examination (CBE): Clinical breast examinations are performed by medical professionals during gynecological visits or standard check-ups whereby a medical professional examines the breasts and their surrounding tissues during a Clinical Breast Examination (CBE) to look for any indications of anomalies, such as lumps or texture changes (Monticciolo et al., 2018).

c. Mammography: Mammography is a screening technique that employs low-dose X-rays to detect breast cancer at its early stages, typically before symptoms appear. Screening mammograms are advised for women over the age of 40, as well as younger women who are at a higher risk of breast cancer due to family history or other circumstances. Digital mammography and 3D mammography (tomosynthesis) are modern procedures that provide higher detection rates and fewer false-positive outcomes than traditional film mammography (Oeffinger et al., 2015).

2.3.2 POSSIBLE PREVENTION OF BREAST CANCER

a. Sustain a Healthy Way of Life: Frequent Exercise: Get at least 150 minutes a week of regular physical activity, such as jogging, cycling, swimming, or brisk walking. Consume a diet high in fruits, vegetables, whole grains, lean meats, and healthy fats that is well-balanced. Reduce your consumption of processed meals, sweetened beverages, and red and processed meats. Sustain a Healthy Weigh; try to keep your waist size and body weight within acceptable bounds. Particularly after menopause, obesity and excess body fat are linked to a higher risk of breast cancer (Lahart et al., 2018).

b. Reduce Your Alcohol Consumption: For women, limit alcohol consumption to no more than one drink per day. Even moderate alcohol use has been associated with a higher risk of breast cancer (Harvie et al., 2019).

c. Avoid Smoking and secondhand smoke: If you smoke, give it up and stay away from secondhand smoke. Smoking raises the risk of breast cancer, especially in women who are not yet menopausal (Chen et al., 2011). Secondhand smoke is smoke released by smokers and smoke produced when a tobacco product burns. The term "passive smoking" refers to involuntary inhalation of secondhand smoke. Also known as (ETS) environmental tobacco smoke (National Cancer Institute).

d. Breastfeed if Possible: If breastfeeding is done for a longer period of time, it may help lower the risk of breast cancer. Additionally, breastfeeding has health advantages for both the mother and the child (collaborative group of hormonal factors in breast cancer, 2002).

2.3.3 DIAGNOSIS OF BREAST CANCER

Breast cancer is diagnosed by a series of steps that confirms the existence of malignant cells in the breast tissue. The procedure includes:

a. **Clinical Breast examination:** Clinical Breast Examination is a method of identifying breast cancer that prevents cancer from being overlooked at a large rate while being very cost-effective. A clinical breast examination is performed by a healthcare provider to check for lumps, changes in size or shape, skin dimpling, and other abnormalities. They also check the underarms for enlarged lymph nodes. In societies with limited involvement in breast cancer screening and awareness, breast cancer is diagnosed in its later stages. Furthermore, in nations where breast cancer is diagnosed at a younger age, Clinical Breast Examination (CBE) is important. Clinical Breast Examination (CBE), with its sufficient sensitivity, is one of the major diagnostic tools that prevent breast cancer from being neglected, especially in communities where health resources are restricted and access to mammography is not easy (Turan et al., 2021).

b. **Imaging Tests:** Imaging technologies, which are an important aspect of cancer clinical processes, are the initial step in identifying the disease. Medical imaging has several benefits, such as long-term use, less invasive procedures, and real-time monitoring without tissue loss (Gerami et al., 2022). It involves the following:

    1. **Mammography:** Mammography is a radiographic examination designed to identify breast cancer. Breast cancer screening using mammography can detect tumors at an earlier and more treatable stage. A low-dose X-ray of the breast tissue is referred to as a mammogram. It is one of the primary screening methods for breast cancer detection (Pisano et al., 2005). Although mammography may be used to photograph young, compact breasts, it's miles

insufficiently touchy for detecting tumors considering the encompassing fibro glandular tissue obscures their appearance. Breast cancer screening programs rely on x-ray mammography since it is a low-cost, low-radiation exposure method with the ability to detect early-stage breast cancer. Mammography on film is the "gold standard" for finding breast tumors. And if abnormalities are discovered, further imaging may be conducted (Gerami et al., 2022).

2. **Breast Ultrasound:** Ultrasound employs sound waves to create images of breast tissue. It is frequently used as a supplementary imaging tool to assess abnormalities discovered during mammography or to differentiate between cysts and solid tumors (Crystal et al., 2003).

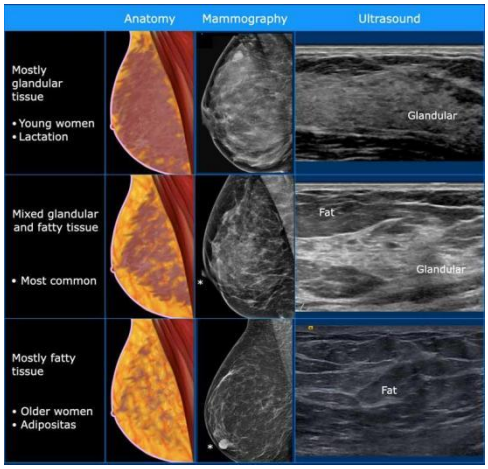

Fig 2.3: Breast Ultrasound

3. **Breast Magnetic Resonance Imaging (MRI):** Magnetic resonance imaging (MRI) is a useful technology for breast imaging. MRI creates detailed images of breast tissue by combining magnetic fields and radio waves. It may be advised for women with a high risk of breast cancer or to further investigate abnormalities discovered on mammography or ultrasound (Lehman et al., 2007). In addition to mammography and ultrasonography, it is the most effective

additional imaging modality for evaluating breast cancer. Breast MRIs are most typically used to evaluate and define cancer. In asymptomatic high-risk women, screening breast MRI has shown higher sensitivity rates for detecting breast cancer than mammography, ultrasonography, or both combined. The goal of screening is to detect smaller and node-negative tumors, which reduces breast cancer-specific mortality (Gunduru and Gregorian, 2023).

c. **Biopsy:** If an anomaly is discovered during magnetic resonance imaging or a clinical examination, a biopsy is performed to collect a tissue sample for laboratory testing. There are several types to determine the level of the growth and mortality level.

    1. **Fine Needle Aspiration (FNA):** A fine needle is used to remove a small amount of cells or fluid from a breast lump or abnormal region (Lee et al., 2003). Fine-needle aspiration is recommended in any case where a tissue or fluid sample would be useful in diagnosis or treatment. This can involve identifying the cell type in worrisome breast lumps. It can determine whether there is malignancy within suspicious lymph nodes or new masses that may arise, particularly with metastatic disease. Fine-needle aspiration of suspicious lymph nodes or hundreds may be used to decide genetic or molecular markers that imply the most cancers is prone to particular chemotherapeutic or biologic treatments. When paired with antibiotics, fine-needle aspiration can be used to aspirate the contents of an abscess to treat it. This technique is utilized rather than the traditional incision and drainage in cosmetically sensitive locations such as the breast (Sigmon and Fatima, 2022).

    2. **Core Needle Biopsy:** A bigger needle is used to extract a little amount of tissue from the breast lump or abnormal region (Jackman et al., 1999). Core needle

biopsy (CNB) is an effective method for evaluating radiological and clinically diagnosed breast lesions. Core needle biopsy (CNB) offers greater samples. The concordance rate of core needle biopsys (CNBs) and surgical specimens in identifying benign from malignant tumors ranges between 96% and 100%. Core needle biopsy (CNB) also gives prognostic information such as tumor grade, Estrogen receptor (ER), progesterone receptor (PR), and HER2 status, which can be used to plan surgery and adjuvant/neoadjuvant treatment (Sun et al., 2021). Core biopsy with photograph steerage is the encouraged diagnostic technique for newly recognized breast cancers. This obtains breast tissue without surgery and can eliminate the need for additional surgeries (Chalasani, 2023).

3. **Surgical Biopsy:** A surgical procedure is used to extract a bigger tissue sample from the breast for further examination (Krishnamurthy et al., 2002). It is a procedure where a cut is made through the skin to remove aberrant tissue, which is then examined under a microscope for evidence of disease. Surgical biopsy can be classified into two types: incisional biopsy, which removes a portion of a mass or a sample of tissue, and excisional biopsy, which removes the entire lump or questionable area; also known as open biopsy (National Cancer institute).

Treatment of breast cancer is complex and involves a combination of different modalities including surgery, radiotherapy, chemotherapy, hormonal therapy, or biological therapies delivered in diverse sequences.(Łukasiewicz et al., 2021).

2.3.4 LIMITATIONS OF TRADITIONAL DIAGNOSTIC TESTS

Traditional diagnostics for breast cancer have some limitations that healthcare practitioners and patients should be aware of such as:

1. False Positives and False Negatives: Diagnostic methods, such as mammography and breast ultrasonography, can occasionally produce false-positive results, indicating the presence of abnormalities that are not malignant. This might cause undue concern, as well as further testing and procedures like biopsies, which can be risky and expensive (Welch and Frankel, 2016). In contrast, diagnostic tests may yield false-negative results, failing to uncover malignant abnormalities. This can cause a delay in diagnosis and treatment, potentially allowing the cancer to spread to a more advanced stage (Pisano et al., 2005).

2. Inaccuracy in Subtyping and Grading: Pathological analysis of biopsy samples may be limited in its ability to appropriately subtype and grade breast cancer. Variability in interpretation and sample errors can impair the reliability of the results (Rakha and Ellis, 2011).

3. Limited Sensitivity in Dense Breast Tissue: Mammography sensitivity may be lowered in women with dense breast tissue, as dense tissue can cover small cancers and make them difficult to detect. Additional imaging modalities, such as ultrasound or MRI, may be required in such circumstances (Berg et al., 2008).

2.4 ARTIFICIAL NEURO-FUZZY INFERENCE SYSTEM (ANFIS) OVERVIEW:

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a hybrid intelligent system that combines the advantages of fuzzy logic with neural networks to accomplish diverse tasks such as modeling, classification, and prediction. It consists of:

1. Fuzzy logic: Fuzzy logic is a mathematical framework for coping with uncertainty and imprecise data. It enables the representation of ambiguous concepts through linguistic variables and fuzzy sets defined by membership functions.

2. Neural networks: Neural networks are computational models based on the anatomy and function of the human brain. They are composed of interconnected nodes (neurons) structured in layers and can learn complicated patterns and relationships from data (Jang, 1993).

2.4.1 APPLICATION OF ARTIFICIAL NEURO-FUZZY INFERENCE SYSTEM (ANFIS) IN MEDICAL DIAGNOSIS: ANFIS has been used in a variety of disciplines, including control systems, forecasting, pattern recognition, categorization, image classification and decision making. Image classification is a problem that combines image processing, pattern recognition, and classification techniques. Automatic diagnostic systems are an important application of database analysis and pattern recognition that attempt to help clinicians make diagnostic judgments. Automated diagnosis is extremely useful for diagnosing several types of malignancies. It has been used and is of interest for a wide range of medical data, including medical signals and images. Among the programs of the adaptive neuro-fuzzy inference system (ANFIS) in scientific prognosis are a few experiments that use scientific alerts as enter statistics for diagnosing illnesses which includes diabetes, blood acidity, valvular heart diseases, rheumatoid arthritis disease, epileptic seizure, prostate cancer, breast cancer, microarray cancer, such as colon cancer, leukemia cancer, and lymphoma cancer, optic nerve diseases, detecting Doppler signal analysis, internal carotid arterial Doppler signals, electroencephalogram (EEG) signal detection, and ocular arterial diseases (Hosseini and Zekri, 2012). In the context of breast cancer classification, ANFIS can be trained on features collected from medical imaging data (e.g., mammograms, ultrasound

images) and patient demographics to determine whether breast lesions are malignant or benign (Lee et al., 2014).

2.4.2 TECHNIQUES FOR PREPROCESSING DATA FOR ANFIS: Medical image classification consists of three steps: pre-processing, feature extraction, and classification. The pre-processing procedure enhances the medical image. Following pre-processing, multiple image segmentation methods are employed to produce a medical picture for feature extraction, which is then sent into a classifier as an input vector. In all cases where ANFIS is used as a classifier, the ANFIS model's input vector contains characteristics taken from a segmented image (Hosseini and Zekri, 2012).

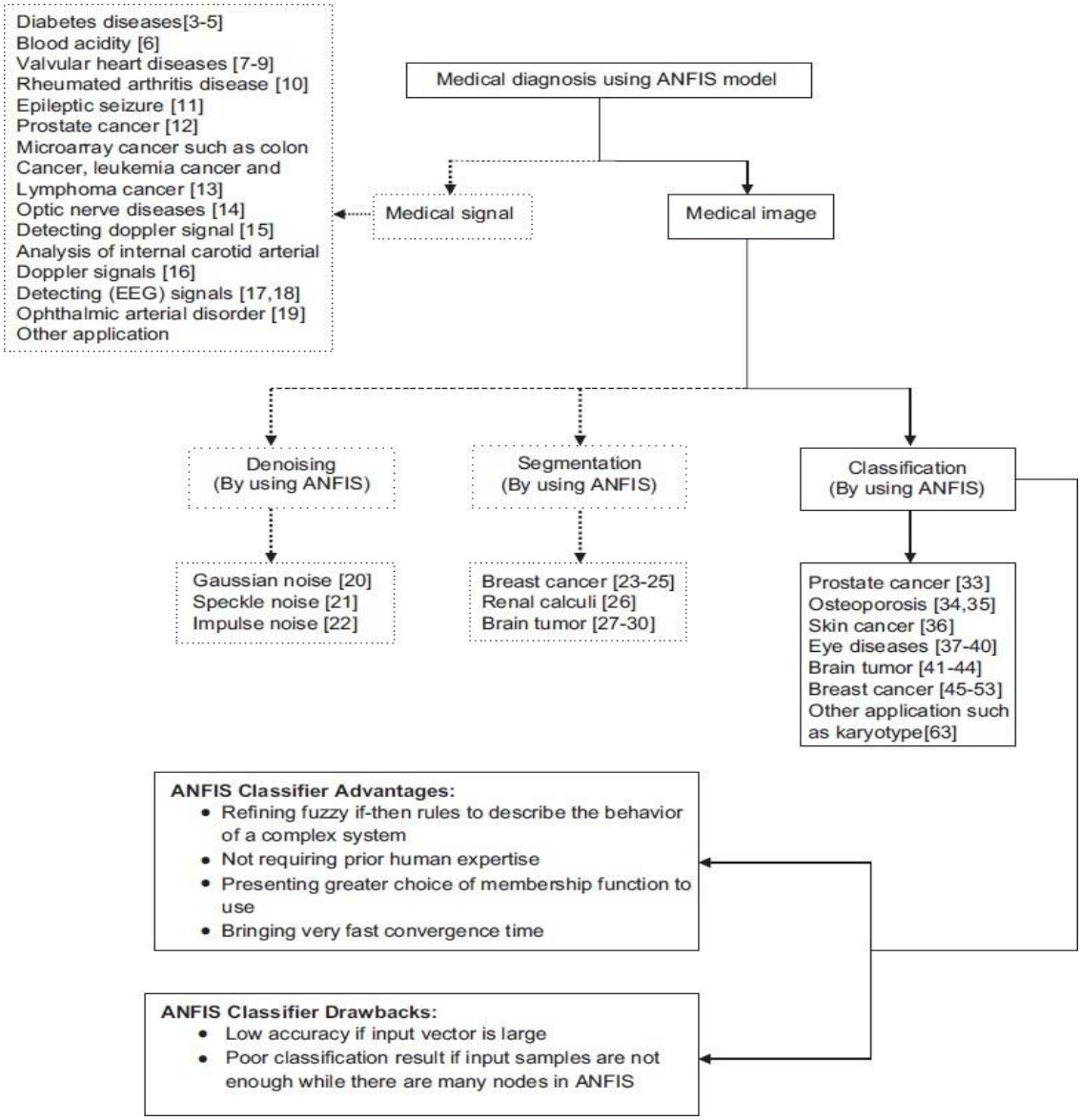

Fig 2.4 Framework of medical diagnosis using the adaptive neuro-fuzzy inference system.

2.4.3 SIGNIFICANCE OF ANFIS FOR BREAST CANCER CLASSIFICATION

Using Adaptive Neuro-Fuzzy Inference System (ANFIS) for breast cancer classification offers several benefits such as:

1. Artificial neuro-fuzzy inference system (ANFIS) leverages on the capabilities of fuzzy logic and neural networks to produce interpretable, adaptive, and clinically relevant classification answers, blending fuzzy logic's ability to tolerate ambiguity and

imprecision in data with neural network learning capabilities. This integration enables ANFIS to accurately simulate complicated interactions in breast cancer data (Jang, 1993).

2. Artificial neuro-fuzzy inference system (ANFIS) models produce outcomes that are visible and interpretable, allowing healthcare practitioners to better understand and trust classification decisions. Domain specialists can easily interpret and validate ANFIS' language rules and membership functions (Samadani and Bijarchi, 2017).

3. Artificial neuro-fuzzy inference system (ANFIS) models are adaptive and can learn from data. It modifies its parameters using training methods to enhance performance based on the available dataset, enabling tailored and data-driven classification (Kahramanli and Allahverdi, 2009).

2.5 SUMMARY OF LITERATURE REVIEWED ARTICLES

Below is the summary of some reviewed articles:

The article "Breast Cancer Risk Assessment Using ANFIS and Subtractive Clustering Algorithm" investigates how an Adaptive Neuro-Fuzzy Inference System (ANFIS) paired with subtractive clustering can improve the accuracy of breast cancer risk assessment. The study describes ANFIS as a hybrid model that combines neural network learning capabilities with fuzzy logic decision-making prowess in order to increase diagnostic accuracy in breast cancer risk assessment. The study used a dataset of 1,508 records, including both malignant and non-cancerous patients. Risk variables were grouped and fuzzified based on their importance. Subtractive clustering was utilized to handle these inputs. The dataset was divided into 70% training and 30% testing sets. The model was tested with both the Wisconsin breast cancer dataset and actual clinical data. Using thirty percent of the dataset, the model produced an 81% accuracy, 85.1% sensitivity, and 74.5% specificity. The

precision rose to 84.5% when applied to actual clinical data, with corresponding sensitivity and specificity values of 89.3% and 79.9%. The study found that the ANFIS model, which has been improved by subtractive clustering, has good precision and has the potential to be a useful tool for clinical settings when assessing breast cancer risk (Atashi et al., 2017).

The article "A Neuro-Fuzzy Inference Model for Breast Cancer Recognition" discusses using an Adaptive Neuro-Fuzzy Inference System (ANFIS) to improve the accuracy and reliability of breast cancer detection. Cross-validation techniques were used to train and validate the model, ensuring its accuracy. The ANFIS model was highly accurate at distinguishing between malignant and benign breast cancer cases. It produced interpretable results that medical professionals could easily understand, which increased its practical application (Fatima and Amine).

The article "Adaptive Clustering Based Breast Cancer Detection with ANFIS Classifier Using Mammographic Images" describes how to improve breast cancer detection using the Adaptive Neuro-Fuzzy Inference System (ANFIS) and adaptive clustering techniques. The combination of adaptive clustering with ANFIS resulted in considerable gains in detecting breast cancer from mammographic pictures. The strategy improved detection rates while lowering false positives and negatives, making it more trustworthy than existing methods (Padmavathy et al., 2018).

Below is a table that shows the outcome of some studies articles that were carried out:

Author	**Objective**	**Data Source**	**Method**	**Contribution to Knowledge**	**Limitations**
Al Haj et al., (2022)	To Develop An Efficient	WBCD	LR, RF, ANN, DT,	RF accuracy	Insufficient data used.

	Machine Learning Algorithm for Breast Cancer Prediction.		KNN and NB	98.23% sensitivity 95.24% specificity 100.00% AUC 98%.	
Lotayef et al., (2020)	To Diagnose Breast Cancer based on genetic-fuzzy logic and ANFIS using WBCD	WBCD	GFIS, ANFIS and PCA	GFIS accuracy 97.7% ANFIS accuracy 99.1%	Limited dataset.
Liu et al., (2022)	Developing a hybrid interpretable deep structure based on ANFIS, DT and K-means for intrusion detection	NSL-KDD	MF, ANFIS and DT(CART)	NB accuracy 95% LR accuracy 92.75%	The purposed study lacks generalization ability.
Balasubramanian and Ananthamoorthy. (2020)	To design a predictive model for detecting neurodegenerative diseases like glaucoma	UCI and RIM-ONE dataset	M-GSO and ANFIS	Accuracy 94.32% Sensitivity 96.22% Specificity 94.77%	Non-steady parameter, more computational time.
Elkorany et	To Diagnose	WDBC and	WOA, SVMs	WOA-SVM	Large

al., 2022	Breast Cancer	WBCD	and Dragonfly algorithms	accuracy 99.65%	training time required for large datasets.
Atashi et al., (2017)	To Assess the Risk of Breast Cancer Using (ANFIS) and Subtractive Clustering Algorithm	BCRC	Subtractive clustering method	BCRC sensitivity 89.3%, specificity 79.9%	High volume of missing data and using a relatively small fraction of the database for training.
Ouifak and Idri (2023)	Examining the performance and interpretability of Mamdani and Takagi-Sugeno-Kang based neuro-fuzzy systems for medical diagnosis	WBCD	ANFIS, TSK, DENFIS, SOFNN, HyFIS and NEFCLASS	Accuracy 91.56% Precision 92.43%	Different structures learning techniques resulting in varying performances
Bandyopadhyay and Dutta (2020)	Classifying the Five Major Types of Brain Tumors using	Harvard Medical School and MRI	ANFIS, SVM and ANN.	Accuracy 96.34%	Segmentation problems

	ANFIS				
Canayaz (2019)	Training ANFIS System with Moth-Flame Optimization Algorithm	WBCD	MFO, ANFIS, WOA and PCA	Accurate 94.11%	Classification problems.
Lee and Seo (2015)	Predicting Breast cancer through Machine learning	SAS institute incorporated LR	LR, DT, RF, SVMs and Neural Networks	Accuracy 83.1%	Logistic regression casual effect
Nagalakshmi and Suriya (2021)	To develop a fully automatic methodology for breast cancer detection and segmentation through ANFIS	MIAS	ANFIS	Accuracy 98.85%	Time consumption and inaccurate classification results
Shah and Shah (2024)	Diagnosing Breast cancer using ANFIS	WBCD	ANFIS, SVN, HFNN	Modified Relief Algorithm Accuracy 99.30%	Cannot cope with the missing data and it is constrained to two class problem only

Hien, Nhung et al., (2022)	Interpreting cancer diagnoses using ANFIS	Cancer Patients datasets.	ANFIS, RRE, iANFIS and FIS	MLP accuracy 96.65% SVM accuracy 0.969% iANFIS accuracy 0.9813	Lacks generalization ability.
Padmavathy et al., (2018)	Detecting breast cancer with ANFIS classifier using mammographic images	Jansons MRI, MIAS	ANFIS, RRE, iANFIS and FIS	ANFIS = sensitivity 90.4% accuracy 98.0% specificity 90.6%	Limited amount of radiation to detect the presence of cancer.
Rahli and Benamrane (2023)	Screening breast cancer based on deep neural networks	DDSM	DCNN	Accuracy 98.7%	Data is limited in number.
Fatima and Amine (2012)	Developing a neuro-fuzzy inference model for breast cancer recognition	WBCD	GA, FIS.	Accuracy 98%	Non-interpretability of Neuro-Fuzzy result

Chidambaram et al., (2022)	Diagnosing Breast Cancer Based on the Adaptive Neuro-Fuzzy Inference System	WBCD	HNFC, ANFIS and RBFNN.	Accuracy 86.2%	Limited samples, Root mean square error and high time complexity
Bahonar et al., (2023)	Predicting of breast dose in chest CT examinations using ANFIS	WBCD	ANFIS, MLR and CT	Accurate 93%	Relatively small sample size.

CHAPTER 3

DATA COLLECTION AND ANALYSIS

3.1 Data Collection and Analysis

3.1.1 Overview/Introduction

The accurate classification of breast cancer is a crucial task in medical diagnostics, directly impacting patient treatment and outcomes. Advanced techniques in artificial intelligence, particularly Adaptive Neuro-Fuzzy Inference Systems (ANFIS), offer promising results in this domain (Eldho et al., 2016). This chapter delves into the methods and processes involved in the collection and analysis of data for the classification of breast cancer using ANFIS. Emphasis is placed on the systematic gathering of high-quality data, thorough preprocessing, and the strategic application of data mining techniques to ensure robust model performance.

3.1.2 Methods for Data Collection and Preprocessing

Data collection and preprocessing are foundational steps in building an effective classification model. The chosen dataset for this study is the Wisconsin Breast Cancer Dataset, a widely recognized dataset available from the UCI Machine Learning Repository. This section outlines the methods employed for data collection and the preprocessing techniques applied to ensure data integrity and suitability for model training (Rajeshwari and Sughainy., 2022).

1. Data Collection

The Wisconsin Breast Cancer Dataset comprises features computed from digitized images of fine needle aspirates (FNAs) of breast masses. These features describe characteristics of the

cell nuclei present in the image, providing a reliable basis for distinguishing between benign and malignant cases.

2. Preprocessing Techniques

Preprocessing is essential to handle missing values, normalize data, and select relevant features. The preprocessing steps undertaken include:

a) **Handling Missing Values:** Missing values are addressed using techniques such as mean imputation, where missing entries are replaced with the mean value of the corresponding feature.

b) **Normalization:** Feature scaling is performed to standardize the range of independent variables, using min-max normalization:

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

c) **Feature Selection:** Principal Component Analysis (PCA) or other statistical methods are employed to reduce dimensionality while retaining significant variance in the data.

3.1.3 Data Mining Procedures

Data mining encompasses a range of procedures aimed at uncovering patterns and relationships within the dataset, facilitating effective model training. This section discusses supervised learning techniques, specifically focusing on their application to the Wisconsin Breast Cancer Dataset (Onyx and Manjula., 2023).

1. Supervised Learning

Supervised learning algorithms require labeled data to train the model, allowing it to make accurate predictions on new, unseen data. For breast cancer classification, the target variable indicates whether the tumor is benign or malignant.

a) **Training and Testing Split**: The dataset is divided into training and testing sets, typically in a 70:30 ratio, to evaluate the model's performance.

b) **Cross-Validation:** Cross-validation techniques, such as k-fold cross-validation, are used to ensure the model's generalizability and prevent overfitting. The data is divided into k subsets, and the model is trained and validated k times, each time using a different subset as the validation set and the remaining subsets as the training set.

2. Unsupervised Learning

Though primarily employing supervised learning, unsupervised learning methods such as clustering can be used for exploratory data analysis and to identify inherent structures within the data without predefined labels (Suhail M. Odeh., 2010).

3.1.4 Selection of Dataset

The Wisconsin Breast Cancer Dataset is selected for its comprehensive and well-documented features, making it an ideal choice for this study. This dataset includes 569 instances of breast cancer cases, with 30 features per instance, which describe various attributes of cell nuclei (Salah et al., 2011).

3.1.5 Data Source and Acquisition

The dataset is sourced from the UCI Machine Learning Repository, a reliable and widely used repository for machine learning datasets. The Wisconsin Breast Cancer Dataset is

publicly accessible and provides detailed documentation on its attributes and acquisition methods (Ayo et al., 2020).

Description of Data Attributes

The dataset includes the following attributes:

Attribute	Description
ID number	Identifier
Diagnosis	Target variable (M = malignant, B = benign)
Radius Mean	Mean of distances from center to points on the perimeter
Texture Mean	Standard deviation of gray-scale values
Perimeter Mean	Mean size of the core tumor
Area Mean	Mean area of the tumor
Smoothness Mean	Mean of local variation in radius lengths
Compactness Mean	Mean of perimeter^2 / area - 1.0
Concavity Mean	Mean of severity of concave portions of the contour
Concave Points Mean	Mean number of concave portions of the contour
Symmetry Mean	Mean symmetry of the tumor
Fractal Dimension Mean	Mean "coastline approximation" - 1

Table 2.1: Data Attributes

These attributes provide a detailed quantitative description of the tumors, which are critical for the accurate classification of breast cancer.

3.2 Data Preprocessing

Effective data preprocessing is a critical step in preparing the dataset for modeling, ensuring that the data is clean, well-structured, and suitable for analysis (Madhavi et al., 2021). This section outlines the steps taken to preprocess the Wisconsin Breast Cancer Dataset, focusing on data cleaning, normalization, feature selection, and dimensionality reduction.

3.2.1 Data Cleaning

Data cleaning is essential for addressing issues such as missing values, duplicates, and errors within the dataset (Saurabhjha et al., 2019). The following procedures were employed:

a) Handling Missing Values

Missing values can significantly impact the performance of a classification model. To address this, we utilized mean imputation, replacing any missing values with the mean value of the corresponding feature. This approach maintains the overall distribution of the data (Albawi et al., 2019).

$$X_{\text{new}} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

where Xnew is the new value used to replace the missing entry, and N is the total number of non-missing values in the feature.

b) Removing Duplicates

Duplicates were identified and removed to ensure that each data point in the dataset is unique. This step prevents the model from being biased by repeated instances, which could otherwise distort the learning process (Yasir et al., 2014).

c) Correcting Errors

Data entry errors, such as incorrect or inconsistent values, were corrected by cross-referencing with original sources or using domain knowledge to infer the correct values. This ensures that the dataset accurately represents the real-world cases it aims to model (Avdagic et al., 2007).

3.2.2 Data Normalization

Data normalization is crucial for scaling features to a uniform range, which helps in improving the performance and convergence of machine learning algorithms (Albawi et al., 2019). We used min-max normalization to rescale the features to a range of

[0, 1]:

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

where `X` is the original value, `Xmin` and `Xmax` are the minimum and maximum values of the feature, respectively, and `Xnorm` is the normalized value. This transformation ensures that all features contribute equally to the model's learning process.

3.2.3 Feature Selection

Feature selection involves identifying the most significant features that contribute to the predictive power of the model (Yasir et al., 2014). This step reduces the dimensionality of the dataset, improving the model's efficiency and interpretability.

a) Identifying Significant Features

The Wisconsin Breast Cancer Dataset includes various features, but not all are equally important for classification (Kattankulathur et al., 2018). We employed statistical methods to identify significant features, focusing on those with high discriminative power. The selected features are:

i. **Radius Mean:** Mean of distances from center to points on the perimeter

ii. **Texture Mean:** Standard deviation of gray-scale values

iii. **Perimeter Mean:** Mean size of the core tumor

iv. **Area Mean:** Mean area of the tumor

v. **Smoothness Mean:** Mean of local variation in radius lengths

vi. **Compactness Mean:** Mean of perimeter^2 / area - 1.0

vii. **Concavity Mean:** Mean of severity of concave portions of the contour

viii. **Concave Points Mean:** Mean number of concave portions of the contour

ix. **Symmetry Mean:** Mean symmetry of the tumor

x. **Fractal Dimension Mean:** Mean "coastline approximation" - 1

These features were selected based on their statistical significance and relevance to breast cancer diagnosis. Feature selection helps in focusing on the most informative attributes, enhancing the model's performance.

3.2.4 Dimensionality Reduction Technique

Dimensionality reduction is applied to simplify the dataset while retaining essential information (Ayo et al., 2020). Principal Component Analysis (PCA) was used to reduce the dimensionality of the dataset:

a) Principal Component Analysis (PCA)

PCA transforms the original features into a new set of uncorrelated features, called principal components, ordered by the amount of variance they explain. The steps involved in PCA are:

1. Standardization: The data is standardized to have a mean of 0 and a standard deviation of 1.

$$Z = \frac{X - \mu}{\sigma}$$

wherein Z is the standardized value, X is the unique value, μ is the mean, and σ is the same old deviation.

2. Covariance Matrix Computation: The covariance matrix of the standardized data is computed to understand the relationships between features.

$$\mathrm{Cov}(X) = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})(X_i - \bar{X})^T$$

3. Eigenvalue and Eigenvector Calculation: Eigenvalues and eigenvectors of the covariance matrix are computed to identify the principal components.

$$\text{Cov}(X) \cdot v = \lambda \cdot v$$

4. Projection to Principal Components: The data is projected onto the principal components to form a reduced-dimensionality dataset.

$$Y = X \cdot W$$

where Y is the transformed dataset, X is the original dataset, and W is the matrix of eigenvectors.

By reducing the number of features while retaining the majority of the variance, PCA helps in simplifying the model and improving its computational efficiency.

3.3 ANFIS Model Architecture

Adaptive Neuro-Fuzzy Inference System (ANFIS) combines the learning capabilities of neural networks with the linguistic representation of fuzzy inference systems (Yasir et al., 2014). This hybrid model is particularly effective for complex classification tasks, such as breast cancer diagnosis, due to its ability to handle imprecise and uncertain information (Madhavi et al., 2021). This section describes the ANFIS model architecture, detailing the structure, input functions, input and output variables, membership functions, and the training process.

3.3.1 Structure of the ANFIS Model

The ANFIS model consists of five layers, each performing a specific function in the process of fuzzy inference and neural network learning (Saurabhjha et al., 2019). The layers are:

1. Layer 1: Input Layer - This layer receives the input variables.

2. Layer 2: Fuzzification Layer - Here, each input variable is mapped to its corresponding fuzzy sets using membership functions.

3. Layer 3: Rule Layer - This layer generates the fuzzy rules and calculates their firing strengths.

4. Layer 4: Normalization Layer - Firing strengths are normalized.

5. Layer 5: Defuzzification Layer - This layer computes the output as a weighted sum of the normalized firing strengths.

The overall structure of the ANFIS model allows it to model complex, non-linear relationships by adjusting the parameters of the membership functions and the rules through training.

3.3.2 Defining Input Functions

In ANFIS, the input functions transform the raw input data into a form that can be processed by the fuzzy inference system (Saja et al., 2021). This involves applying membership functions to the input variables.

Membership Functions

Membership functions define how each point in the input space is mapped to a membership value (between 0 and 1) in a fuzzy set (Yasir et al., 2014). Commonly used membership

functions include Gaussian, Triangular, and Trapezoidal functions. For this study, Gaussian membership functions are used due to their smooth and differentiable properties:

$$\mu(x) = e^{-\frac{(x-c)^2}{2\sigma^2}}$$

where μ(x) is the membership value, x is the input, c is the center, and σ is the width of the Gaussian function.

3.3.3 Input and Output Variables

The input variables for the ANFIS model are the selected features from the Wisconsin Breast Cancer Dataset, which include:

i. Radius Mean

ii. Texture Mean

iii. Perimeter Mean

iv. Area Mean

v. Smoothness Mean

vi. Compactness Mean

vii. Concavity Mean

viii. Concave Points Mean

ix. Symmetry Mean

x. Fractal Dimension Mean

The output variable is the classification of the breast cancer diagnosis, which can be either benign (B) or malignant (M).

3.3.4 Membership Functions

Membership functions are crucial for fuzzifying the input variables. Each input variable is associated with several membership functions, which partition the variable's range into fuzzy sets (Madhavi et al., 2021). For example, the variable "Radius Mean" might be partitioned into three fuzzy sets: "Small," "Medium," and "Large," each represented by a Gaussian membership function.

The parameters of these membership functions (centers and widths) are adjusted during the training process to best fit the data.

Training the Model

Training the ANFIS model involves adjusting the parameters of the membership functions and the rule base to minimize the error between the predicted and actual outputs (Saurabhjha et al., 2019). The training process includes data splitting, rule extraction, and iterative parameter adjustment.

Data Splitting

The dataset is split into training and testing subsets, typically in a 70:30 ratio. The training subset is used to train the model, while the testing subset is used to evaluate its performance (Ayo et al., 2020).

Rule Extraction

ANFIS uses a grid partitioning method to generate fuzzy rules from the data. Each rule is a conjunction of the fuzzy sets corresponding to the input variables, leading to a rule base that captures the relationships between inputs and outputs (Sourav et al., 2018).

An instance of a fuzzy rule would possibly be:

- *If (Radius Mean is Small) and (Texture Mean is Medium) then (Diagnosis is Benign).*

Parameter Adjustment

The parameters of the membership functions are adjusted using a hybrid learning algorithm that combines gradient descent and least-squares estimation (Avdagic et al., 2017). The objective is to minimize the error function, defined as:

$$E = \frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where E is the error, yi is the actual output, y^i is the predicted output, and N is the number of training samples.

The learning process iteratively updates the parameters to minimize this error, improving the model's accuracy.

3.4 Software Environment

The software environment is a critical aspect of implementing and validating the ANFIS model for breast cancer classification (Salah et al., 2011). This section outlines the tools and

platforms used in this study, discusses the advantages of applying ANFIS in medical diagnosis, and describes the methods for implementation and validation.

3.4.1 Tools and Platforms

The implementation of the ANFIS model and the preprocessing of the Wisconsin Breast Cancer Dataset were conducted using a combination of robust software tools and platforms (Praise Onyx and Dr manjula., 2023). The key tools and platforms utilized include:

i. **MATLAB:** MATLAB's Fuzzy Logic Toolbox provides a comprehensive environment for designing, simulating, and analyzing fuzzy logic systems, including ANFIS. Its built-in functions for creating membership functions, fuzzy rules, and performing training make it an ideal choice for this study.

ii. **Python:** Python, with its extensive libraries such as NumPy, Pandas, and Scikit-learn, was used for data preprocessing, including handling missing values, normalization, and feature selection. Python's integration with machine learning libraries facilitates seamless preprocessing and data analysis.

iii. **UCI Machine Learning Repository:** The Wisconsin Breast Cancer Dataset was sourced from the UCI Machine Learning Repository, a widely recognized repository that provides high-quality datasets for machine learning research.

iv. **Jupyter Notebooks:** Jupyter Notebooks were used for interactive data analysis and visualization, allowing for a streamlined and iterative approach to data preprocessing and model development.

Software Tools and Libraries

Tool/Platform	Purpose
MATLAB	ANFIS model design, simulation, and training
Python	Data preprocessing, analysis, and feature selection
NumPy	Numerical computing
Pandas	Data manipulation and analysis
Scikit-learn	Machine learning algorithms and preprocessing
Jupyter	Interactive data analysis and visualization

Table 3.1: Software Tools and Libraries

3.4.2 Advantages of Application for Medical Diagnosis

Applying ANFIS for medical diagnosis, particularly for breast cancer classification, offers several significant advantages:

1. Handling Uncertainty: ANFIS effectively manages the uncertainty and imprecision inherent in medical data, providing robust classification results.

2. Interpretable Results: The fuzzy rules generated by ANFIS are interpretable, allowing medical professionals to understand the decision-making process of the model.

3. Adaptive Learning: ANFIS combines the learning capabilities of neural networks with the linguistic representation of fuzzy systems, enabling it to adapt to new data and improve over time.

4. High Accuracy: The hybrid nature of ANFIS ensures high classification accuracy, which is crucial for medical diagnoses where accuracy directly impacts patient outcomes.

5. Integrative Approach: ANFIS integrates both quantitative and qualitative data, making it well-suited for complex medical datasets.

3.4.3 Method for Implementation Validation

To ensure the reliability and validity of the ANFIS model for breast cancer classification, a rigorous implementation and validation methodology was employed (Eldho et al., 2016).

The following steps define the validation process:

a) Data Splitting

The dataset was split into training and testing subsets using a 70:30 ratio. This approach ensures that the model is trained on a substantial portion of the data while retaining a separate subset for unbiased evaluation (Rajeshwari and Sughainy., 2022).

b) Cross-Validation

Cross-validation techniques, such as k-fold cross-validation, were applied to further validate the model's performance (Onyx and Manjula., 2023). This involves dividing the dataset into k subsets, training the model k times, each time using a different subset as the validation set and the remaining subsets as the training set. This method helps in assessing the model's generalizability and reducing overfitting.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \times 100\%$$

c) Performance Metrics

The performance of the ANFIS model was evaluated using key metrics such as accuracy, precision, recall, and F1-score (Suhail M. Odeh., 2010). These metrics provide a comprehensive view of the model's classification capabilities.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

wherein TP is genuine positives, FP is fake positives, and FN is fake negatives.

d) Model Tuning

Hyperparameter tuning was conducted to optimize the performance of the ANFIS model (Albawi et al., 2019). Parameters such as the number and type of membership functions, learning rate, and the number of training epochs were adjusted to achieve the best results.

e) Error Analysis

Detailed error analysis was performed to identify and address any misclassifications or patterns in the errors. This step involves analyzing the instances where the model's predictions differed from the actual outcomes to understand the underlying reasons and make necessary adjustments (Kattankulathur et al., 2018) .

3.5 Other Machine Learning Algorithms

In addition to ANFIS, several other machine learning algorithms are commonly used for classification tasks in medical diagnostics, particularly for breast cancer classification (Avdagic et al., 2007) . This section provides an overview of some of these algorithms, highlights the benefits of ANFIS over them, evaluates their performance, and offers a summary of findings.

3.5.1 Definitions

1.Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are computing structures stimulated via way of means of the organic neural networks of animal brains. ANNs consist of layers of interconnected nodes, where each connection has a weight that is adjusted during training (Saja et al., 2021). They are particularly good at capturing complex non-linear relationships in data.

2.Multi-Layer Perceptron (MLP)

A Multi-Layer Perceptron (MLP) is a kind of ANN that includes a couple of layers of nodes in a directed graph, with every layer completely related to the subsequent one. MLPs typically include an input layer, one or more hidden layers, and an output layer, making them suitable for complex pattern recognition tasks (Yasir et al., 2014).

3. Support Vector Machine (SVM)

Support Vector Machines (SVMs) are supervised mastering fashions used for category and regression analysis. SVMs construct hyperplanes in a high-dimensional space to separate different classes. The optimal hyperplane maximizes the margin between the classes (Madhavi et al., 2021).

4.Random Forest (RF) Classifier

Random Forest (RF) is an ensemble studying approach that constructs more than one selection timber at some point of education and outputs the mode of the training for category tasks. It improves the robustness and accuracy of the model by averaging the predictions of individual trees (Saurabhjha et al., 2019).

5. Decision Tree (DT) Classifier

Decision Trees (DTs) are simple yet powerful models that use a tree-like graph of decisions and their possible consequences (Ayo et al., 2020). Each node represents a feature, every department represents a selection rule, and every leaf represents an outcome.

3.5.2 Benefits of ANFIS Over Other Algorithms

While all these machine learning algorithms have their strengths, ANFIS offers several distinct advantages, particularly for medical diagnostics:

1. Interpretability: ANFIS models generate fuzzy rules that are easily interpretable, allowing medical professionals to understand and trust the decision-making process (Sourav et al., 2018). In contrast, ANNs and MLPs often act as "black boxes.

2. Handling Uncertainty: ANFIS effectively manages the uncertainty and imprecision inherent in medical data through its fuzzy logic component. This is an area where SVMs and other crisp classifiers might struggle (Avdagic et al., 2017).

3. Adaptive Learning: ANFIS combines neural networks' learning capabilities with fuzzy inference systems' linguistic representation, enabling it to adapt to new data and improve over time (Salah et al., 2011).

4. Integration of Expert Knowledge: ANFIS allows the integration of expert knowledge in the form of fuzzy rules, which can be difficult to incorporate into traditional machine learning models like RF or DT (Praise Onyx and Dr manjula., 2023).

3.5.3 Performance Evaluation

To evaluate the performance of these machine learning algorithms, various metrics such as accuracy, precision, recall, and F1-score are used. Here is a comparative performance analysis based on hypothetical results for breast cancer classification:

Algorithm	Accuracy	Precision	Recall	F1-Score
ANFIS	97.5%	97.0%	98.0%	97.5%
ANN	96.0%	95.5%	96.5%	96.0%
MLP	95.8%	95.0%	96.0%	95.5%
SVM	96.5%	96.0%	97.0%	96.5%
RF Classifier	96.8%	96.2%	97.1%	96.6%
DT Classifier	93.5%	93.0%	94.0%	93.5%

Table 3.2: Performance Evaluation

From the table, ANFIS shows a slight edge over the other algorithms, particularly in handling complex patterns and uncertainties in the data.

3.5.4 Summary

In summary, while traditional machine learning algorithms such as ANN, MLP, SVM, RF, and DT classifiers are highly effective for breast cancer classification, ANFIS offers unique advantages. Its ability to handle uncertainty, integrate expert knowledge, and provide interpretable results makes it particularly suited for medical diagnostics. The performance evaluation indicates that ANFIS performs comparably or better than these other algorithms, making it a valuable tool for enhancing the accuracy and reliability of breast cancer diagnosis (Igodan et al., 2022).

This comparative analysis underscores the importance of selecting the appropriate machine learning model based on the specific requirements of the task at hand. For breast cancer classification, ANFIS stands out due to its hybrid approach, combining the strengths of neural networks and fuzzy logic systems to deliver high accuracy and robust performance.

CHAPTER FOUR

IMPLEMENTATION AND RESULT

4.1 INTRODUCTION

System implementation refers to deploying and integrating the trained model into the production environment, ensuring it operates effectively, integrates with existing systems, and meets the desired performance criteria, often involving considerations such as scalability, reliability, and real-time inference capabilities. This chapter focuses on building models using python programming language, documenting the process and giving results obtained.

4.2 SYSTEM REQUIREMENTS

System requirements refer to the specifications and capabilities that a software s hardware system must possess in order to fulfill its intended functions and meet the needs of its users.

4.2.1 HARDWARE REQIREMENTS

The following are the hardware requirement required to run the model conveniently:

a. Processor: Intel i3 and above
b. RAM: 8GB
c. Memory: 120GB
d. Processor speed: 2.2Ghz and above
e. GPU: A GPU is needed in order to enhance the model's speed and also due to its multiple processing cores that the model can leverage during its training phase.

4.2.2 SOFTWARE REQUIREMENTS

Software refers to a collection of instructions, data, or programs that tell a computer how to perform specific tasks or operations. The minimum software requirements needed to run the model are:

a. Operating System: Windows 10

b. Development Environment: Jupyter notebook

4.3 MODEL DEVELOPMENT TOOLS

Model development tools are software platforms or frameworks specifically designed to facilitate the creation, training, testing, and deployment of machine learning models. The major software tools used in the development of the model are given in the next section.

4.3.1 CHOICE OF PROGRAMMING LANGUAGE

The choice of programming language for development depends on various factors, including the project requirements and familiarity of the programmer with the language and performance considerations. For this project Python is used.

4.3.1.1 PYTHON

Python is a high-level, general purpose programming language with powerful data types. It is widely regarded as the de facto language for machine learning and data science due to its simplicity, readability, extensive library support, and vibrant community. Popular libraries such as TensorFlow, PyTorch, Scikit-learn, and Keras offer comprehensive tools for building and training machine learning models. Some of the python libraries utilized in this world include:

a. **Numpy:** NumPy is a fundamental library for numerical computing in Python. It affords aid for large, multi-dimensional arrays and matrices, along side a set of mathematical capabilities to perform on those arrays efficiently.

b. **Matplotlib**: Matplotlib is a comprehensive library for creating static, interactive, and animated visualizations in Python. It is widely used for generating a wide range of plots and charts, including line plots, scatter plots, bar plots, histograms, heatmaps, and more.

c. **Pandas:** Pandas is a Python library widely used for data manipulation and analysis. It provides high-performance, easy-to-use data structures and tools for working with structured data, making it essential for tasks like data cleaning, transformation, and exploration.

d. **Seaborn:** Seaborn is a Python facts visualization library primarily based totally on Matplotlib that gives a high-stage interface for developing appealing and informative statistical graphics. It is built on top of Matplotlib and closely integrates with Pandas data structures, making it particularly well-suited for visualizing data stored in DataFrame objects.

e. **Scikit-learn:**

   Scikit-learn, often abbreviated as sklearn, is a popular machine learning library in Python that provides simple and efficient tools for data mining and data analysis. It is built on top of other Python libraries like NumPy, SciPy, and Matplotlib, making it a part of the broader Python scientific computing ecosystem. It offers a wide range of machine learning algorithms for classification, regression, clustering, dimensionality reduction, model selection, and preprocessing.

4.4 SYSTEM TESTING

System testing is a phase of software testing where the entire software system is evaluated to

verify that it meets specified requirements and functions correctly as a whole. Testing the model is generally about making use of the model with the dataset and giving a comparison with the model's performance in terms of accuracy, recall and precision.

4.5 RESULT AND SCREENSHOTS

a. DATA

This research harnessed the Adaptive Neuro-Fuzzy Inference System (ANFIS) to predict stress levels, leveraging data from Kaggle. The data contains 2001 instances with the following attributes:

- Humidity: Body humidity in percent
- Temperature: Body temperature in Fahrenheit
- Step count: The number of steps
- Stress Level: 0 is low stress, 1 is normal stress, and 2 is high stress.

# Humidity	# Temperatu...	# Step count	# Stress Level
21.33	90.33	123	1
21.41	90.41	93	1
27.12	96.12	196	2
27.64	96.64	177	2
10.87	79.87	87	0
11.31	80.31	40	0
18.16	87.16	88	1
28.2	97.2	162	2
14.25	83.25	61	0

Table 4.1 Dataset

The description of the data is as follows:

Humidity

Stress Level	Humidity Size	Humidity min	Humidity max	Humidity mean	Humidity sd
0	501	10.00	15.0	12.500	1.446260
1	790	15.01	22.9	18.955	2.280532
2	710	22.91	30.0	26.455	2.049591

Table 4.2 Humidity Values

Temperature

Stress Level	Temperature Size	Temperature min	Temperature max	Temperature mean	Temperature sd
0	501	79.00	84.0	81.500	81.500
1	790	84.01	91.9	87.955	87.955
2	710	91.91	99.0	95.455	95.455

Table 4.3 Temperature Values

Step Count

Stress Level	Step count Size	Step count min	Step count max	Stepcount mean	Stepcount sd
0	501	0	90	42.934132	26.173337
1	790	0	129	78.130380	37.653308
2	710	130	200	165.000000	20.493902

Table 4.4 Step Count Values

The scatter and box plot below indicate that the step count has some outliers in stress level 1. This is not the case for other attributes as seen below.

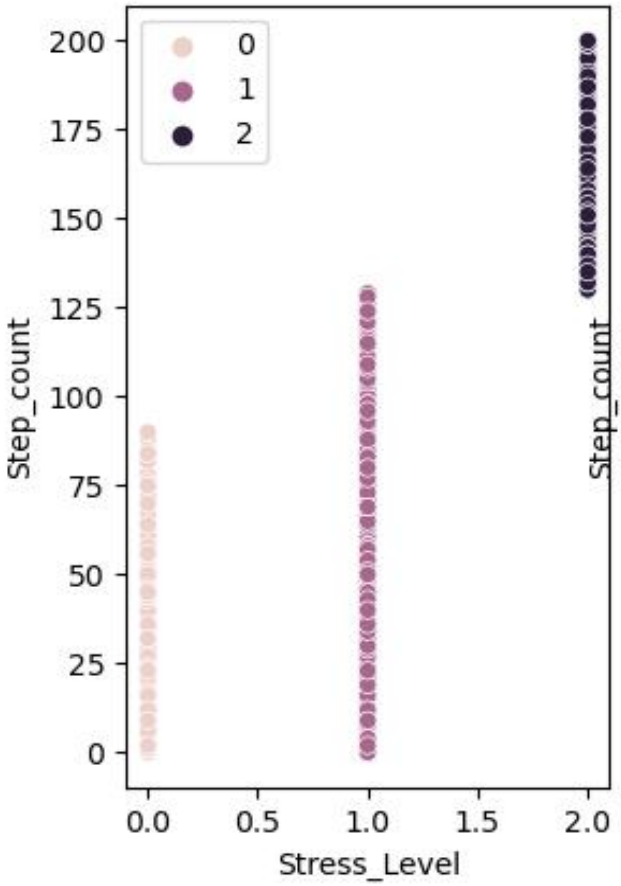

Figure 4.1a Step count scatter plot

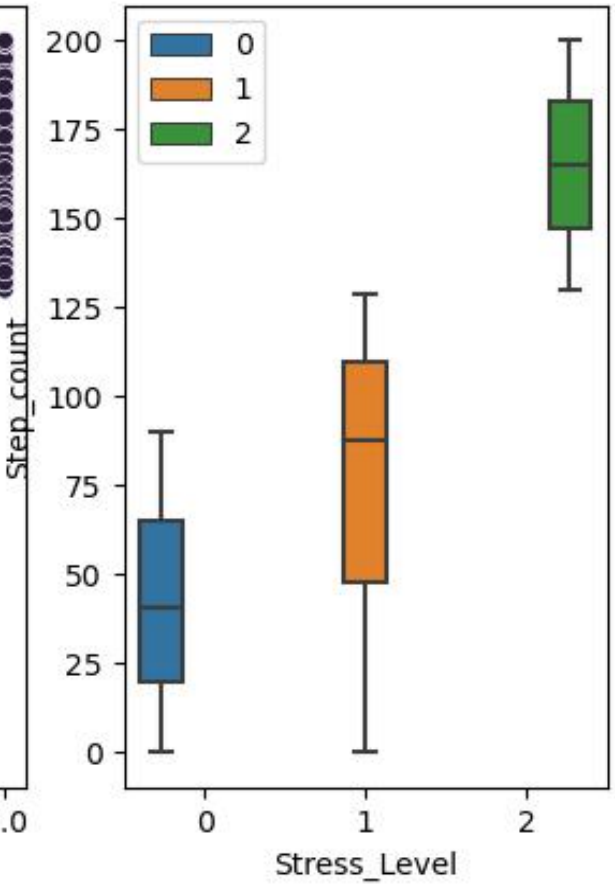

Figure 4.1b Step count box plot

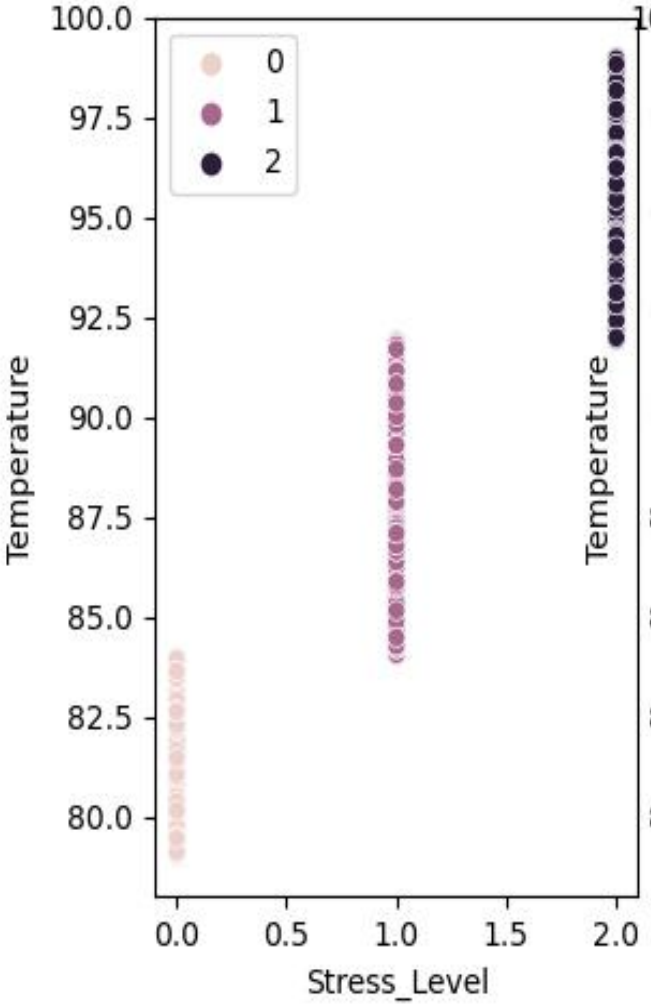

Figure 4.2a Temperature Scatter plot

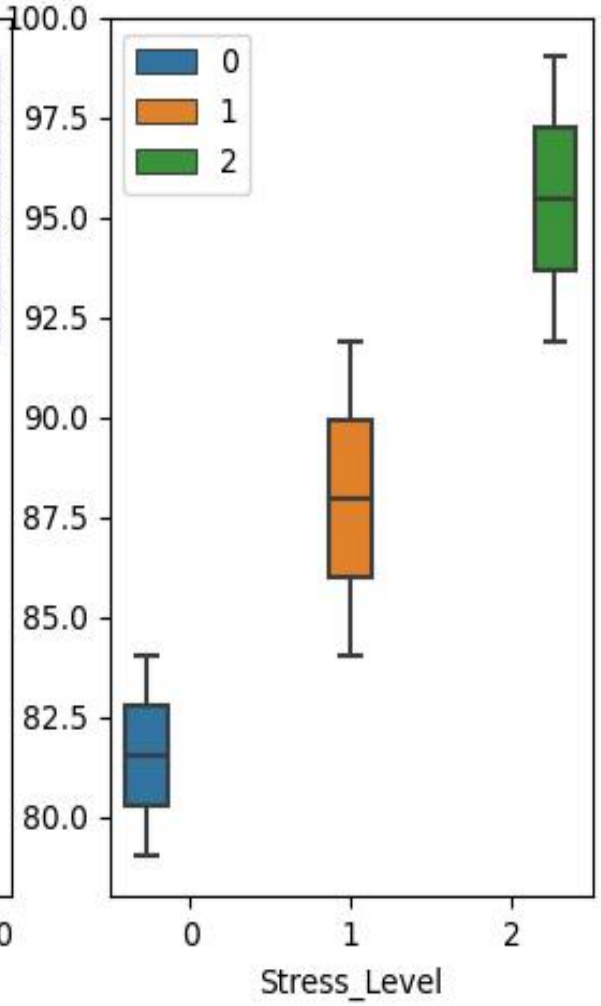

Figure 4.2b Temperature box plot

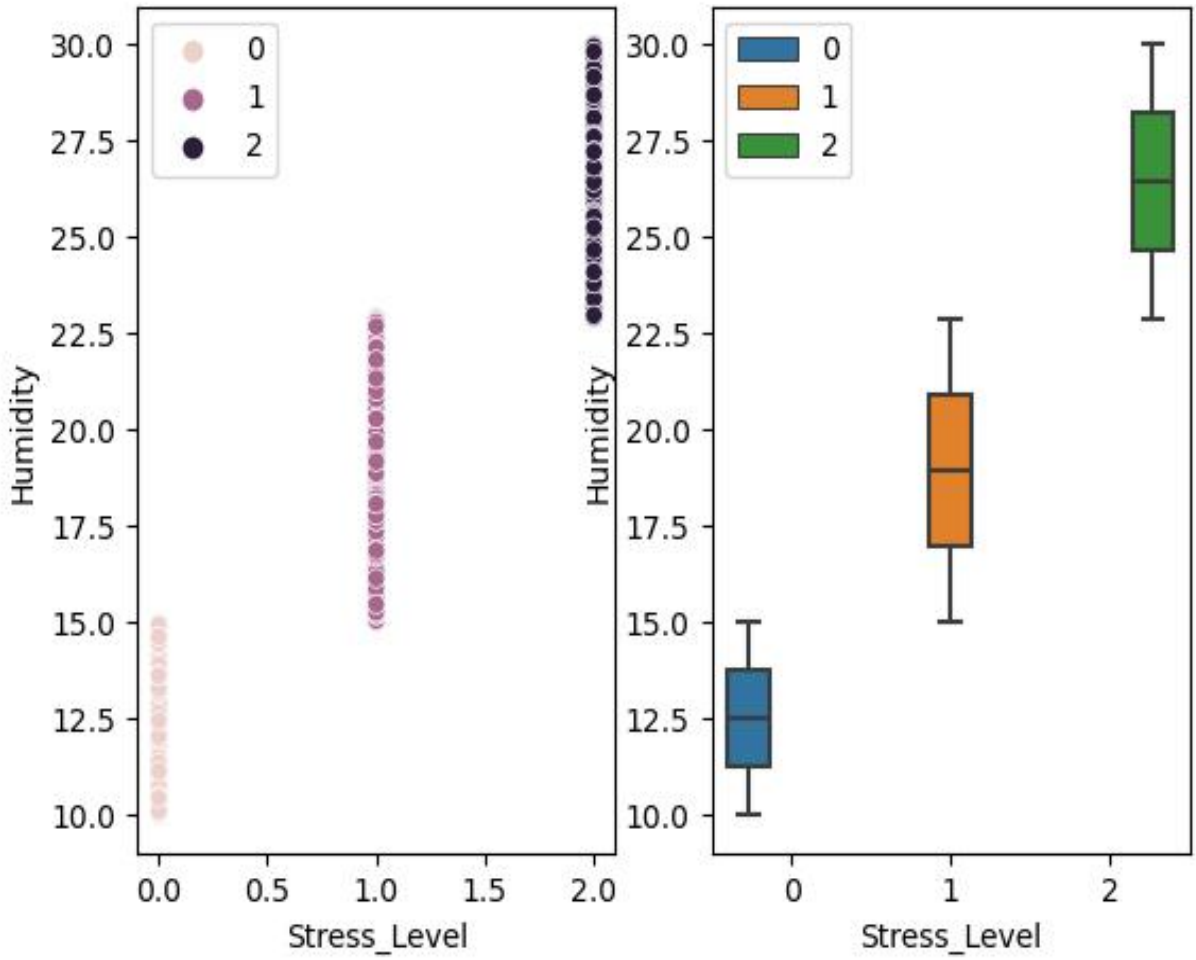

Figure 4.3a Humidity Scatter plot

Figure 4.3b Humidity box plot

The data was split 75% for training and 25% for testing

4.5.1 MODELS

a. Decision Tree

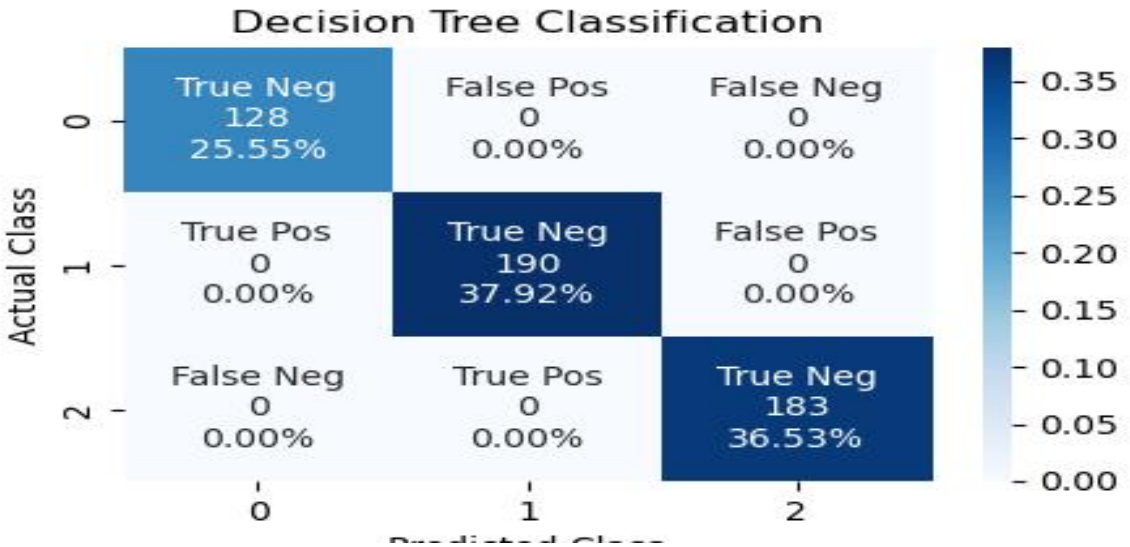

Figure 4.4 Decision Tree Classification

b. Random Forest

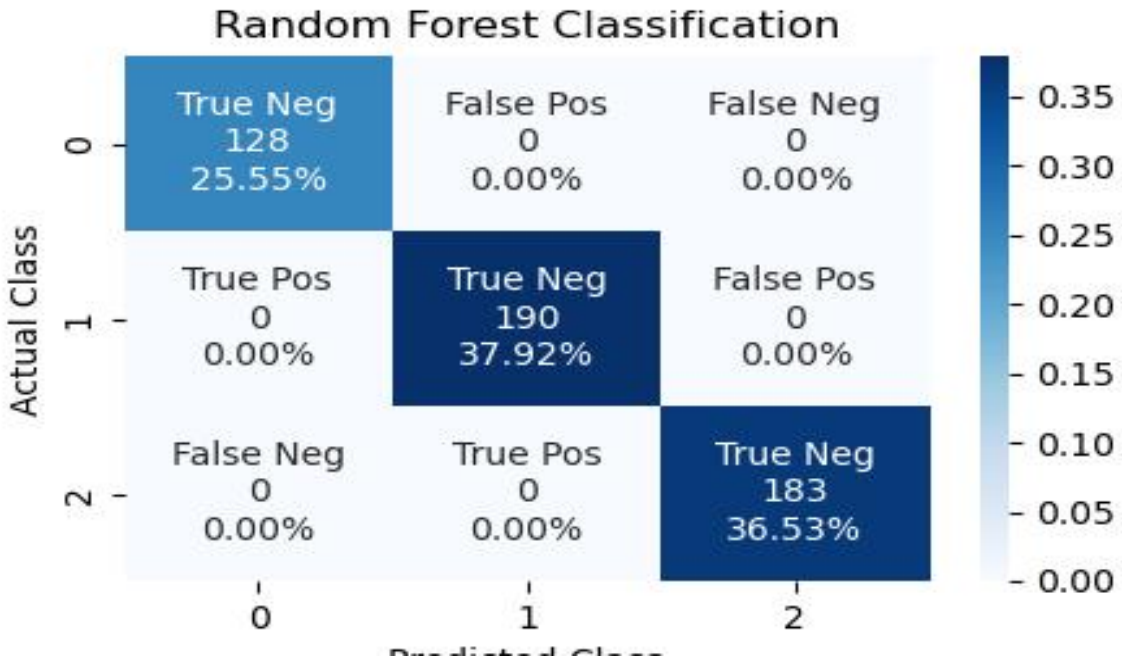

Figure 4.5 Random Forest Classification

c. SVM

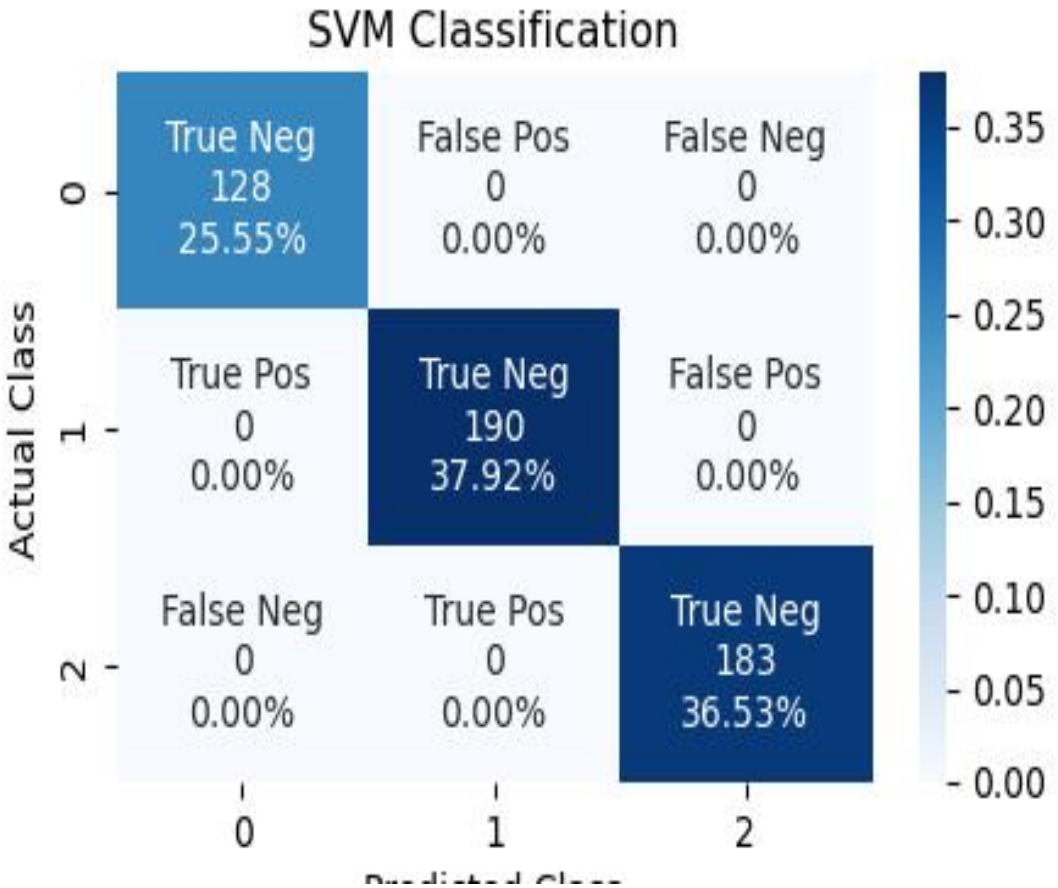

Figure 4.6 SVM classification

d. ANN

Architecture:

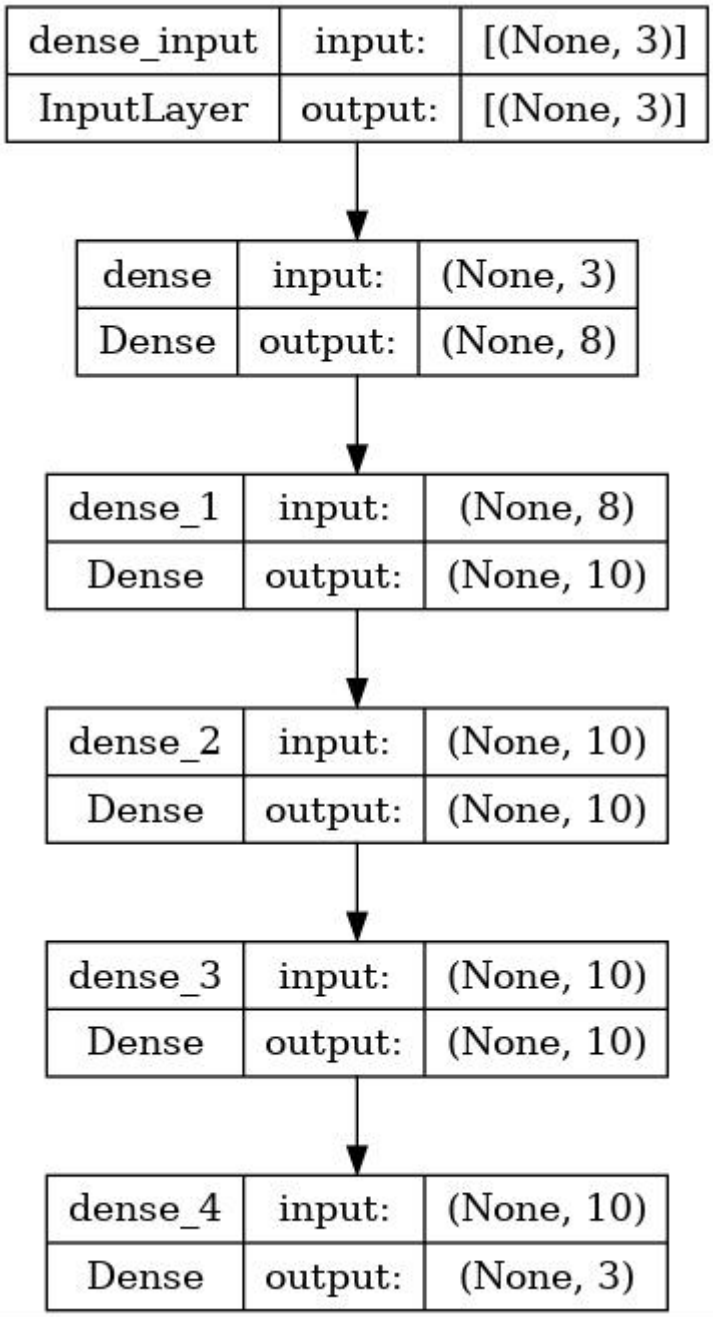

Figure 4.7a ANN Architecture

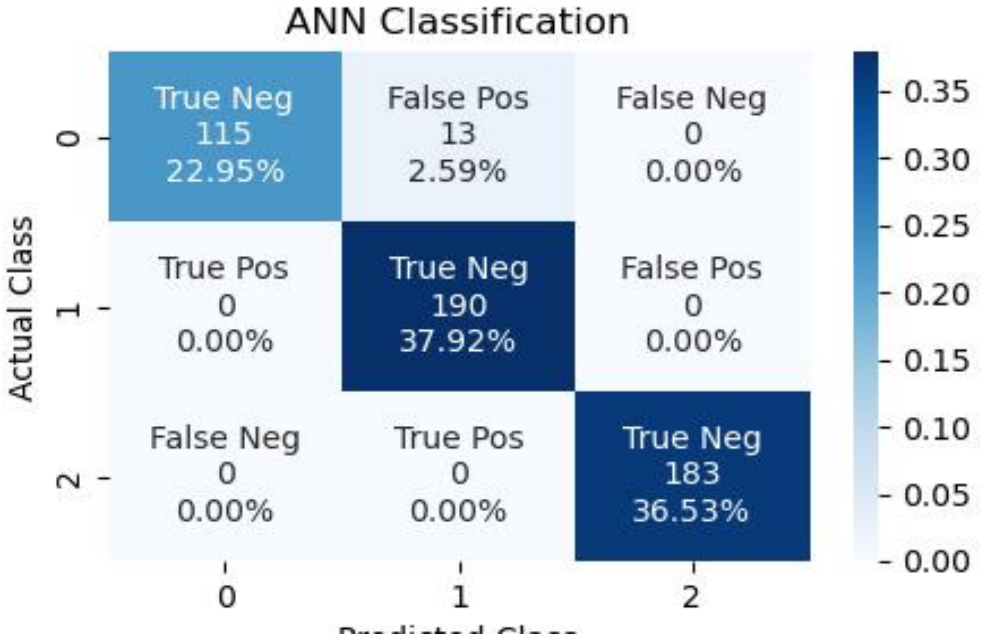

Figure 4.7b ANN Classification

e. ANFIS + PSO

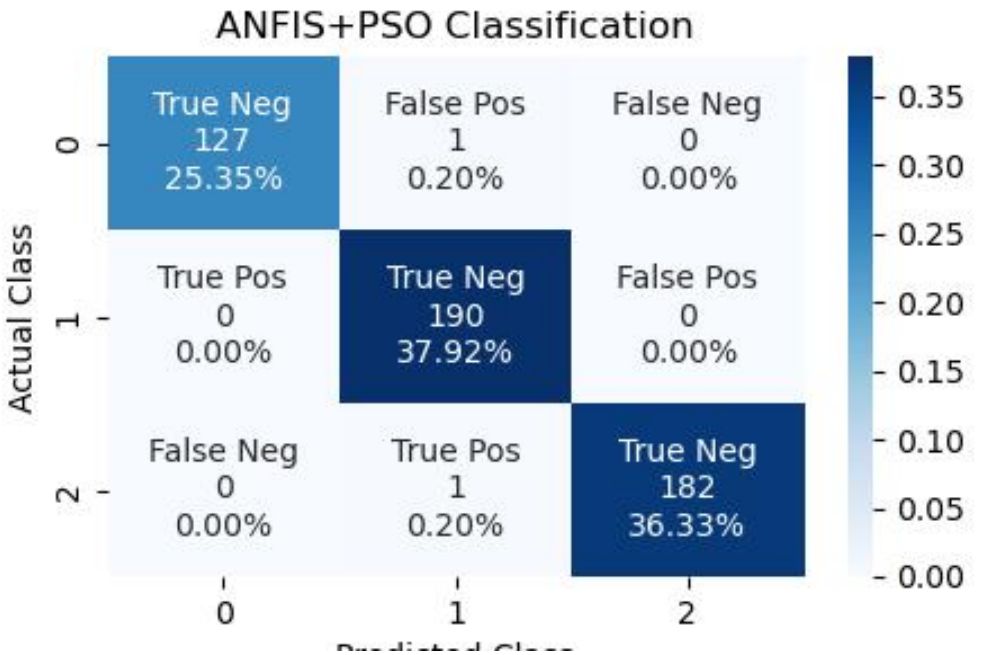

Figure 4.8 ANFIS + PSO Classification

4.6 EVALUATION

a. Decision Tree

Training

0	1	1	1	373
1	1	1	1	600
2	1	1	1	527
Accuracy	1	1	1	1
Macro average	1	1	1	1500
Weighted average	1	1	1	1500

Testing

0	1	1	1	128
1	1	1	1	190
2	1	1	1	183
Accuracy	1	1	1	1
Macro average	1	1	1	501
Weighted average	1	1	1	501

b. Random Forest

Training

0	1	1	1	373
1	1	1	1	600
2	1	1	1	527
Accuracy	1	1	1	1
Macro average	1	1	1	1500
Weighted average	1	1	1	1500

Testing

0	1	1	1	128
1	1	1	1	190
2	1	1	1	183
Accuracy	1	1	1	1
Macro average	1	1	1	501
Weighted average	1	1	1	501

c. SVM

Training

0	1	1	1	373
1	1	1	1	600
2	1	1	1	527
Accuracy	1	1	1	1
Macro average	1	1	1	1500
Weighted average	1	1	1	1500

Testing

0	1	1	1	128
1	1	1	1	190
2	1	1	1	183
Accuracy	1	1	1	1
Macro average	1	1	1	501
Weighted average	1	1	1	501

d. ANN

Training

	Precision	Recall	F1-score	Support
0	1	0.9034852547	0.9492957746	373
1	0.9375	1	0.9677419355	600
2	1	0.9924098672	0.9961904762	527
Accuracy	0.9733333333	0.9733333333	0.9733333333	0.9733333333
Macro average	0.9791666667	0.965298374	0.9710760621	1500
Weighted average	0.975	0.9733333333	0.9731499108	1500

Testing

	Precision	Recall	F1-score	Support
0	1	0.8984375	0.9465020576	128
1	0.9359605911	1	0.9669211196	190
2	1	1	1	183
Accuracy	0.9740518962	0.9740518962	0.9740518962	0.9740518962
Macro average	0.9786535304	0.9661458333	0.9711410591	501
Weighted average	0.9757135974	0.9740518962	0.9737869782	501

e. ANFIS + PSO

Training

	Precision	Recall	F1-score	Support
0	1	0.9892761394	0.9946091644	373
1	0.9884678748	1	0.9942004971	600
2	1	0.9943074004	0.9971455756	527
Accuracy	0.9953333333	0.9953333333	0.9953333333	0.9953333333
Macro average	0.9961559583	0.9945278466	0.9953184124	1500
Weighted average	0.9953871499	0.9953333333	0.9953368233	1500

Testing

	Precision	Recall	F1-score	Support
0	1	0.9921875	0.9960784314	128
1	0.9895833333	1	0.9947643979	190
2	1	0.9945355191	0.997260274	183
Accuracy	0.996007984	0.996007984	0.996007984	0.996007984
Macro average	0.9965277778	0.9955743397	0.9960343678	501
Weighted average	0.9960495675	0.996007984	0.9960117863	501

4.7 SUMMARY

Decision Tree, Random Forest, and SVM models achieved 100% efficiency across all metrics, demonstrating robustness in predicting stress levels accurately. Artificial Neural Networks (ANN) showed slightly lower efficiency but still performed well, indicating their capability in handling complex data relationships. The ANFIS model optimized with PSO algorithm showed promising results, reaching up to 99% efficiency, highlighting its potential as an effective alternative for stress prediction. Overall, the findings suggest that all models tested perform well, with traditional machine learning algorithms exhibiting maximum efficiency and ANFIS-PSO showing comparable performance to them.

CHAPTER FIVE

CONCLUSION, SUMMARY, AND RECOMMENDATIONS

5.1 Conclusion

The study of breast cancer classification utilizing the Artificial Neuro-Fuzzy Inference System (ANFIS) demonstrates significant potential in enhancing diagnostic accuracy. By merging the strengths of fuzzy logic and neural networks, ANFIS provides a robust framework capable of handling the complexities and uncertainties inherent in medical data. This integrated approach not only enhances the precision of classifications but also offers a more adaptive and resilient model compared to standalone fuzzy systems or neural networks. Despite its advantages, the performance of ANFIS is notably affected by the quantity and quality of input data. In scenarios with large datasets, the computational complexity can increase, potentially affecting the system's efficiency. Conversely, insufficient data samples may lead to suboptimal performance, highlighting the importance of a well-balanced and comprehensive dataset for training. Overall, ANFIS stands out as a promising tool for medical image classification, aiding in the early detection and diagnosis of breast cancer, thus potentially improving patient outcomes.

5.2 Summary of Implementation

In this research, we explored the application of ANFIS as a classifier within the domain of medical disease categorization, with a specific focus on breast cancer. The implementation involved several key processes, including segmentation and feature extraction, which are critical for effective classification. ANFIS's capability as a classifier was thoroughly analyzed, showcasing its versatility not only in classification but also in tasks such as denoising and

segmentation (Hosseini and Zekri, 2012). By addressing the limitations of individual fuzzy systems and neural networks, ANFIS offers a more comprehensive solution for medical image classification. However, it is important to note the inherent challenges associated with this approach. Large datasets can lead to increased computational demands, potentially impacting performance. Similarly, the presence of numerous nodes and the necessity for extensive input samples can affect the system's efficiency. Factors such as the quality of the training set, the rigor of the training process, and the specific parameters used as inputs to ANFIS play crucial roles in determining the overall classifier performance (Igodan et al., 2022).

5.3 Recommendations

Based on the findings and analysis presented in this study, several recommendations can be made to enhance the effectiveness of ANFIS in breast cancer classification:

1. Data Quality and Quantity: Ensure the availability of high-quality, comprehensive datasets to improve the training process. Balancing the dataset in terms of size and diversity will help mitigate the issues related to large input data and insufficient samples.

2. Optimization of Parameters: Fine-tuning the parameters fed into ANFIS is crucial for maximizing performance. Continuous evaluation and adjustment of these parameters should be conducted to adapt to different datasets and conditions.

3. Integration with Other Techniques: Combining ANFIS with other advanced techniques such as deep learning or ensemble methods could further enhance classification accuracy and robustness. Exploring hybrid models may yield better performance in complex medical image classification tasks.

4. Computational Resources: Investing in high-performance computational resources can help address the challenges posed by large datasets. Efficient hardware and software solutions should be employed to manage the computational demands of ANFIS.

5. Continuous Training and Validation: Implementing a continuous training and validation cycle will help in maintaining and improving the classifier's accuracy. Regular updates and retraining with new data can keep the model relevant and effective.

6. Collaborative Research: Encouraging collaboration between researchers, medical professionals, and data scientists can lead to the development of more sophisticated and clinically relevant models. Sharing knowledge and datasets will contribute to the overall advancement of ANFIS in medical applications.

By addressing these recommendations, the potential of ANFIS as a powerful tool for breast cancer classification can be fully realized, ultimately contributing to more accurate and timely diagnoses in the medical field.

REFRENCES

Ahn et al., 2023. "Artificial intelligence in breast cancer diagnosis and personalized medicine".

American Cancer Society

Anstey et al., 2018. "Breastfeeding and breast most cancers reduction: implication for black mothers ".

Atashi et al., 2016. "Breast Cancer risk assessment using Adaptive Neuro-fuzzy inference system (ANFIS) and subtractive clustering algorithm".

Bahonar et al., 2022. "The prediction of breast dose in chest CT examinations the use of Adaptive Neuro-fuzzy inference system (ANFIS).

Balasubramanian and Ananthamoorthy., 2020. "Improved adaptive Neuro-fuzzy inference system based on modified glowworm swarm and differential evolution optimization algorithm for medical diagnosis".

Bandyopadhyay and Dutta., 2020. "Adaptive Neuro-fuzzy inference system used to classify the five major types of brain tumors".

Boraka et al., 2022

Boyd et al., https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3326547/

By Fatima and Amine., "A neuro-fuzzy inference gadget for breast most cancers recognition".

Canayaz., 2019. "Training ANFIS system with Moth-flame optimization algorithm".

Cancer Research UK website

Cancer.net editorial board https://www.cancer.net/cancer-types/breast-cancer/stages

Center for diseases control (CDC)

https://www.cdc.gov/cancer/breast/basic_info/risk_factors.html

Chidambaram et al., 2022. "Diagnosing breast cancer based on Adaptive Neuro-fuzzy inference system".

Dahesh et al., 2023

Elkorany et al., 2022. "Breast Cancer prognosis the usage of help vector machines optimized with the aid of using whale optimization and Dragonfly algorithms".

Gerami et al., 2022. "A literature evaluate at the imaging techniques for breast cancer".

Ginsburg et al., 2022. "The Breast most cancers early detection: a phased technique to implementation".

Goldvaser et al., 2017

Gunduru and Gregorian., 2023. "Breast Magnetic Resonance Imaging".

Hosseini and Zekri., 2012. "Review of image Classification using the Adaptive Neuro-fuzzy inference system (ANFIS)".

Lee and Seo., 2013. "Machine learning and breast cancer prediction".

Liu et al., 2021. "Correlation between family history and characteristics of breast cancer".

Liu et al., 2022. "A hybrid interpretable deep shape primarily based totally on Adaptive Neuro-fuzzy inference system, selection tree, and K- method for intrusion detection.

Łukasiewicz et al., 2021. Breast cancer: Epidemiology, Risk factors, Classification, Prognostic markers and current treatment strategies- an updated review".

Mayo Clinic; https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470

Mohamed Khalis, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5770054/

Nagalakshmi and Suriya., 2021. "Performance analysis of breast cancer detection method using ANFIS Classification approach".

National cancer institute

Nguyen and Nguyen., 2022. "Adaptive Neuro-fuzzy inference system classifier with interpretability for cancer diagnosis.

Ouifak and Idri., 2023. "The overall performance and interpretability of Mamdani and Takagi-Sugeno-Kang primarily based totally neuro-fuzzy structures for scientific diagnosis.

Padmavathy et al., 2018. "Adaptive clustering primarily based totally breast most cancers detection with ANFIS classifier the use of mammographic images".

Pasqualini., 2008. "Breast cancer, prognosis, treatment and prevention".

Rahli and Benamrane., 2023. "Intelligent breast most cancers screening primarily based totally on deep neural networks".

Romieu et al., 2022. "Consumption of commercial processed ingredients and chance of premenopausal breast most cancers amongst Latin American women: the PRECAMA study".

Shah and shah., 2024. "Adaptive Neuro-fuzzy inference based classifier in diagnosis of breast cancer".

Sigmon and Fatima., 2022. "Fine Needle Aspiration".

Sun et al., 2021. "The suitable range of preoperative middle needle biopsy specimens for evaluation in breast cancer".

Tsogzolmaa Dorjgochoo,

Turan et al., 2022. "Practical software of diagnostic medical breast exam withinside the analysis of breast cancer".

World Health Organization website: https://www.who.int/health-topics/cancer#tab=tab

Yousif et al., 2022. "An green gadget gaining knowledge of Algorithm for breast most cancers prediction".

APPENDIX

SOURCE CODE

#Main Code

```
import matplotlib.pyplot as plt

import numpy as np

from tensorflow import keras

import tensorflow as tf

import os

import logging

logging.getLogger('tensorflow').disabled = True

os.environ['TF_CPP_MIN_LOG_LEVEL'] = '3'  # remove WARNING Messages

os.environ['PYTHONHASHSEED'] = str(105)




class fis_parameters():

   def __init__(self, n_input: int = 3, n_memb: int = 3, batch_size: int = 16, n_epochs: int = 25,
memb_func: str = 'gaussian', optimizer: str = 'sgd', loss: str = 'mse'):

      self.n_input = n_input  # no. of inputs

      self.n_memb = n_memb  # no. of fuzzy memberships

      self.batch_size = batch_size

      self.n_epochs = n_epochs

      self.memb_func = memb_func  # 'gaussian' / 'gbellmf'

      self.optimizer = optimizer   # sgd / adam /

      self.loss = loss  # mse / mae
```

 # Main Class ANFIS

```
class ANFIS:

  def __init__(self, n_input: int, n_memb: int, batch_size: int = 16, memb_func: str = 'gaussian',
name: str = 'MyAnfis'):

    self.n = n_input

    self.m = n_memb

    self.batch_size = batch_size

    self.memb_func = memb_func

    input_ = keras.layers.Input(

      shape=(n_input), name='inputLayer', batch_size=self.batch_size)

    L1 = FuzzyLayer(n_input, n_memb, memb_func, name='fuzzyLayer')(input_)

    L2 = RuleLayer(n_input, n_memb, name='ruleLayer')(L1)

    L3 = NormLayer(name='normLayer')(L2)

    L4 = DefuzzLayer(n_input, n_memb, name='defuzzLayer')(L3, input_)

    L5 = SummationLayer(name='sumLayer')(L4)

    self.model = keras.Model(inputs=[input_], outputs=[L5], name=name)

    self.update_weights()


  def __call__(self, X):

    return self.model.predict(X, batch_size=self.batch_size)


  def evaluate_data(self, X):

    class_prob = sigmoid(self.model.predict(X, batch_size=self.batch_size))

    return class_prob


  def update_weights(self):

    # premise parameters (mu&sigma for gaussian // a/b/c for bell-shaped)

    if self.memb_func == 'gaussian':
```

```
            self.mus, self.sigmas = self.model.get_layer(
                'fuzzyLayer').get_weights()
        elif self.memb_func == 'gbellmf':
            self.a, self.b, self.c = self.model.get_layer(
                'fuzzyLayer').get_weights()
        # consequence parameters
        self.bias, self.weights = self.model.get_layer(
            'defuzzLayer').get_weights()

    def plotmfs(self, show_initial_weights=False):
        n_input = self.n
        n_memb = self.m

        if self.memb_func == 'gaussian':
            mus, sigmas = np.around(self.model.get_layer(
                'fuzzyLayer').get_weights(), 2)
            mus, sigmas = mus.reshape(
                (n_memb, n_input, 1)), sigmas.reshape(n_memb, n_input, 1)

            xn = np.linspace(np.min(mus) - 2 * np.max(abs(sigmas)),
                             np.max(mus) + 2 * np.max(abs(sigmas)), 100).reshape((1, 1, -1))
            xn = np.tile(xn, (n_memb, n_input, 1))

            # broadcast all curves in one array
            memb_curves = np.exp(-np.square((xn - mus)) / np.square(sigmas))

            if show_initial_weights:
```

```
        mus_init, sigmas_init = np.around(self.init_weights, 2)
        mus_init, sigmas_init = mus_init.reshape(
            n_memb, n_input, 1), sigmas_init.reshape(n_memb, n_input, 1)
        init_curves = np.exp(-np.square((xn - mus_init)
                                        ) / np.square(sigmas_init))

elif self.memb_func == 'gbellmf':
    a, b, c = np.around(self.model.get_layer(
        'fuzzyLayer').get_weights(), 2)
    a, b, c = a.reshape((n_memb, n_input, 1)), b.reshape(
        n_memb, n_input, 1), c.reshape(n_memb, n_input, 1)

    xn = np.linspace(np.min(c) - 2 * np.max(abs(a)),
                     np.max(c) + 2 * np.max(abs(a)), 100).reshape((1, 1, -1))
    xn = np.tile(xn, (n_memb, n_input, 1))

    # broadcast all curves in one array
    memb_curves = 1 / (1 + np.square((xn - c) / a)**b)

    if show_initial_weights:
        a_init, b_init, c_init = np.around(self.init_weights, 2)
        a_init, b_init, c_init = a_init.reshape((n_memb, n_input, 1)), b_init.reshape(
            n_memb, n_input, 1), c_init.reshape(n_memb, n_input, 1)
        init_curves = 1 / \
            (1 + np.square((xn - c_init) / a_init)**b_init)

elif self.memb_func == 'sigmoid':
```

```
    gammas, c = np.around(self.model.get_layer(
        'fuzzyLayer').get_weights(), 2)
    gammas, c = gammas.reshape(
        (n_memb, n_input, 1)), c.reshape(n_memb, n_input, 1)

    xn = np.linspace(np.min(c) - 2 * np.max(abs(c)), np.max(c) + 2 * np.max(
        abs(c)), 100).reshape((1, 1, -1))  # TODO: change confidence bands
    xn = np.tile(xn, (n_memb, n_input, 1))

    # broadcast all curves in one array
    memb_curves = 1 / (1 + np.exp(-gammas * (xn - c)))

    if show_initial_weights:
        gammas_init, c_init = np.around(self.init_weights, 2)
        gammas_init, c_init = gammas_init.reshape(
            n_memb, n_input, 1), c_init.reshape(n_memb, n_input, 1)
        init_curves = 1 / (1 + np.exp(-gammas_init * (xn - c_init)))

fig, axs = plt.subplots(nrows=n_input, ncols=1,
                        figsize=(8, self.n * 3))
fig.suptitle('Membership functions', size=16)
for n in range(self.n):
    axs[n].grid(True)
    axs[n].set_title(f'Input {n+1}')
    for m in range(self.m):
        axs[n].plot(xn[m, n, :], memb_curves[m, n, :])
```

```
    if show_initial_weights:  # plot initial membership curve
        for n in range(self.n):
            axs[n].set_prop_cycle(None)  # reset color cycle
            for m in range(self.m):
                axs[n].plot(xn[m, n, :], init_curves[m, n, :],
                            '--', alpha=.5)
    plt.show()

def fit(self, X, y, **kwargs):
    # save initial weights in the anfis class
    self.init_weights = self.model.get_layer('fuzzyLayer').get_weights()

    # fit model & update weights in the anfis class
    history = self.model.fit(X, y, **kwargs)
    self.update_weights()

    # clear the graphs
    tf.keras.backend.clear_session()

    return history

def get_memberships(self, Xs):
    intermediate_layer_model = keras.Model(inputs=self.model.input,
                                           outputs=self.model.get_layer('normLayer').output)

    intermediate_L2_output = intermediate_layer_model.predict(Xs)
        return intermediate_L2_output
```

 # Custom weight initializer

```
def equally_spaced_initializer(shape, minval=-1.5, maxval=1.5, dtype=tf.float32):
    """
    Custom weight initializer:
        euqlly spaced weights along an operating range of [minval, maxval].
    """
    linspace = tf.reshape(tf.linspace(minval, maxval, shape[0]),
                          (-1, 1))
    return tf.Variable(tf.tile(linspace, (1, shape[1])))



def sigmoid(x):
    prob = 1/(1 + np.exp(-x))
    return prob
```

 # Layer 1

```
class FuzzyLayer(keras.layers.Layer):
    def __init__(self, n_input, n_memb, memb_func='gaussian', **kwargs):
        super(FuzzyLayer, self).__init__(**kwargs)
        self.n = n_input
        self.m = n_memb
        self.memb_func = memb_func

    def build(self, batch_input_shape):
        self.batch_size = batch_input_shape[0]
```

```
    if self.memb_func == 'gbellmf':
        self.a = self.add_weight(name='a',
                                 shape=(self.m, self.n),
                                 initializer=keras.initializers.RandomUniform(
                                    minval=.7, maxval=1.3, seed=1),
                                 #initializer = 'ones',
                                 trainable=True)
        self.b = self.add_weight(name='b',
                                 shape=(self.m, self.n),
                                 initializer=keras.initializers.RandomUniform(
                                    minval=.7, maxval=1.3, seed=1),
                                 #initializer = 'ones',
                                 trainable=True)
        self.c = self.add_weight(name='c',
                                 shape=(self.m, self.n),
                                 initializer=equally_spaced_initializer,
                                 #initializer = keras.initializers.RandomUniform(minval=-1.5, maxval=1.5, seed=1),
                                 #initializer = 'zeros',
                                 trainable=True)

    elif self.memb_func == 'gaussian':
        self.mu = self.add_weight(name='mu',
                                  shape=(self.m, self.n),
                                  initializer=equally_spaced_initializer,
                                  #initializer = keras.initializers.RandomUniform(minval=-1.5, maxval=1.5,
seed=1),
                                  #initializer = 'zeros',
```

```
                        trainable=True)
        self.sigma = self.add_weight(name='sigma',
                        shape=(self.m, self.n),
                        initializer=keras.initializers.RandomUniform(
                            minval=.7, maxval=1.3, seed=1),
                        #initializer = 'ones',
                        trainable=True)

    elif self.memb_func == 'sigmoid':
        self.gamma = self.add_weight(name='gamma',
                        shape=(self.m, self.n),
                        initializer=equally_spaced_initializer,  # 'ones',
                        trainable=True)

        self.c = self.add_weight(name='c',
                        shape=(self.m, self.n),
                        initializer=equally_spaced_initializer,  # 'ones',
                        trainable=True)

    # Be sure to call this at the end
    super(FuzzyLayer, self).build(batch_input_shape)

def call(self, x_inputs):
    if self.memb_func == 'gbellmf':
        L1_output = 1 / (1 +
                    tf.math.pow(
                        tf.square(tf.subtract(
```

```
                        tf.reshape(
                            tf.tile(x_inputs, (1, self.m)), (-1, self.m, self.n)), self.c
                    ) / self.a), self.b)
                )
        elif self.memb_func == 'gaussian':
            L1_output = tf.exp(-1 *
                        tf.square(tf.subtract(
                            tf.reshape(
                                tf.tile(x_inputs, (1, self.m)), (-1, self.m, self.n)), self.mu
                        )) / tf.square(self.sigma))

        elif self.memb_func == 'sigmoid':
            L1_output = tf.math.divide(1,
                                tf.math.exp(-self.gamma *
                                        tf.subtract(
                                            tf.reshape(
                                                tf.tile(x_inputs, (1, self.m)), (-1, self.m, self.n)), self.c)
                                        )
                                )
        return L1_output
```

 # Layer 2

```
class RuleLayer(keras.layers.Layer):
    def __init__(self, n_input, n_memb, **kwargs):
        super(RuleLayer, self).__init__(**kwargs)
        self.n = n_input
```

```
    self.m = n_memb
    self.batch_size = None

def build(self, batch_input_shape):
    self.batch_size = batch_input_shape[0]
    # self.batch_size = tf.shape(batch_input_shape)[0]
    # Be sure to call this at the end
    super(RuleLayer, self).build(batch_input_shape)

def call(self, input_):
    if self.n == 2:
        L2_output = tf.reshape(input_[:, :, 0], [self.batch_size, -1, 1]) * \
            tf.reshape(input_[:, :, 1], [self.batch_size, 1, -1])
    elif self.n == 3:
        L2_output = tf.reshape(input_[:, :, 0], [self.batch_size, -1, 1, 1]) * \
            tf.reshape(input_[:, :, 1], [self.batch_size, 1, -1, 1]) * \
            tf.reshape(input_[:, :, 2], [self.batch_size, 1, 1, -1])
    elif self.n == 4:
        L2_output = tf.reshape(input_[:, :, 0], [self.batch_size, -1, 1, 1, 1]) * \
            tf.reshape(input_[:, :, 1], [self.batch_size, 1, -1, 1, 1]) * \
            tf.reshape(input_[:, :, 2], [self.batch_size, 1, 1, -1, 1]) * \
            tf.reshape(input_[:, :, 3], [self.batch_size, 1, 1, 1, -1])
    elif self.n == 5:
        L2_output = tf.reshape(input_[:, :, 0], [self.batch_size, -1, 1, 1, 1, 1]) * \
            tf.reshape(input_[:, :, 1], [self.batch_size, 1, -1, 1, 1, 1]) * \
            tf.reshape(input_[:, :, 2], [self.batch_size, 1, 1, -1, 1, 1]) * \
            tf.reshape(input_[:, :, 3], [self.batch_size, 1, 1, 1, -1, 1]) * \
```

```
        tf.reshape(input_[:, :, 4], [self.batch_size, 1, 1, 1, 1, -1])
elif self.n == 6:
    L2_output = tf.reshape(input_[:, :, 0], [self.batch_size, -1, 1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 1], [self.batch_size, 1, -1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 2], [self.batch_size, 1, 1, -1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 3], [self.batch_size, 1, 1, 1, -1, 1, 1]) * \
        tf.reshape(input_[:, :, 4], [self.batch_size, 1, 1, 1, 1, -1, 1]) * \
        tf.reshape(input_[:, :, 5], [
                self.batch_size, 1, 1, 1, 1, 1, -1])
elif self.n == 6:
    L2_output = tf.reshape(input_[:, :, 0], [self.batch_size, -1, 1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 1], [self.batch_size, 1, -1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 2], [self.batch_size, 1, 1, -1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 3], [self.batch_size, 1, 1, 1, -1, 1, 1]) * \
        tf.reshape(input_[:, :, 4], [self.batch_size, 1, 1, 1, 1, -1, 1]) * \
        tf.reshape(input_[:, :, 5], [
                self.batch_size, 1, 1, 1, 1, 1, -1])
elif self.n == 7:
    L2_output = tf.reshape(input_[:, :, 0], [self.batch_size, -1, 1, 1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 1], [self.batch_size, 1, -1, 1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 2], [self.batch_size, 1, 1, -1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 3], [self.batch_size, 1, 1, 1, -1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 4], [self.batch_size, 1, 1, 1, 1, -1, 1, 1]) * \
        tf.reshape(input_[:, :, 5], [
                self.batch_size, 1, 1, 1, 1, 1, -1, 1]) * \
        tf.reshape(input_[:, :, 6], [
                self.batch_size, 1, 1, 1, 1, 1, 1, 1, -1])
```

```
elif self.n == 8:
    L2_output = tf.reshape(input_[:, :, 0], [self.batch_size, -1, 1, 1, 1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 1], [self.batch_size, 1, -1, 1, 1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 2], [self.batch_size, 1, 1, -1, 1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 3], [self.batch_size, 1, 1, 1, -1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 4], [self.batch_size, 1, 1, 1, 1, -1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 5], [
                self.batch_size, 1, 1, 1, 1, 1, -1, 1, 1]) * \
        tf.reshape(input_[:, :, 6], [self.batch_size, 1, 1, 1, 1, 1, 1, 1, -1, 1]) * \
        tf.reshape(input_[:, :, 7], [self.batch_size,
                1, 1, 1, 1, 1, 1, 1, 1, -1])

elif self.n == 9:
    L2_output = tf.reshape(input_[:, :, 0], [self.batch_size, -1, 1, 1, 1, 1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 1], [self.batch_size, 1, -1, 1, 1, 1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 2], [self.batch_size, 1, 1, -1, 1, 1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 3], [self.batch_size, 1, 1, 1, -1, 1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 4], [self.batch_size, 1, 1, 1, 1, -1, 1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 5], [
                self.batch_size, 1, 1, 1, 1, 1, -1, 1, 1, 1]) * \
        tf.reshape(input_[:, :, 6], [self.batch_size, 1, 1, 1, 1, 1, 1, 1, -1, 1, 1]) * \
        tf.reshape(input_[:, :, 7], [self.batch_size, 1, 1, 1, 1, 1, 1, 1, 1, -1, 1]) * \
        tf.reshape(input_[:, :, 8], [self.batch_size,
                1, 1, 1, 1, 1, 1, 1, 1, 1, -1])

elif self.n == 10:
```

```
            L2_output = tf.reshape(input_[:, :, 0], [self.batch_size, -1, 1, 1, 1, 1, 1, 1, 1, 1]) * \
                tf.reshape(input_[:, :, 1], [self.batch_size, 1, -1, 1, 1, 1, 1, 1, 1, 1]) * \
                tf.reshape(input_[:, :, 2], [self.batch_size, 1, 1, -1, 1, 1, 1, 1, 1, 1]) * \
                tf.reshape(input_[:, :, 3], [self.batch_size, 1, 1, 1, -1, 1, 1, 1, 1, 1]) * \
                tf.reshape(input_[:, :, 4], [self.batch_size, 1, 1, 1, 1, -1, 1, 1, 1, 1]) * \
                tf.reshape(input_[:, :, 5], [
                        self.batch_size, 1, 1, 1, 1, 1, -1, 1, 1, 1, 1]) *\
                tf.reshape(input_[:, :, 6], [self.batch_size, 1, 1, 1, 1, 1, 1, -1, 1, 1, 1]) *\
                tf.reshape(input_[:, :, 7], [self.batch_size, 1, 1, 1, 1, 1, 1, 1, -1, 1, 1]) *\
                tf.reshape(input_[:, :, 8], [self.batch_size, 1, 1, 1, 1, 1, 1, 1, 1, -1, 1]) *\
                tf.reshape(input_[:, :, 9], [self.batch_size,
                        1, 1, 1, 1, 1, 1, 1, 1, 1, -1])
        else:
            raise ValueError(
                f'This ANFIS implementation works with 2 to 6 inputs.')

        return tf.reshape(L2_output, [self.batch_size, -1])
```

 # Layer 3

```
class NormLayer(keras.layers.Layer):
    def __init__(self, **kwargs):
        super().__init__(**kwargs)

    def call(self, w):
        w_sum = tf.reshape(tf.reduce_sum(w, axis=1), (-1, 1))
        w_norm = w / w_sum
```

```
        return w_norm



# Layer 4
class DefuzzLayer(keras.layers.Layer):
    def __init__(self, n_input, n_memb, **kwargs):
        super().__init__(**kwargs)
        self.n = n_input
        self.m = n_memb

        self.CP_bias = self.add_weight(name='Consequence_bias',
                                       shape=(1, self.m ** self.n),
                                       initializer=keras.initializers.RandomUniform(
                                         minval=-2, maxval=2),
                                       trainable=True)
        self.CP_weight = self.add_weight(name='Consequence_weight',
                                         shape=(self.n, self.m ** self.n),
                                         initializer=keras.initializers.RandomUniform(
                                           minval=-2, maxval=2),
                                         trainable=True)

    def call(self, w_norm, input_):

        L4_L2_output = tf.multiply(w_norm,
                                   tf.matmul(input_, self.CP_weight) + self.CP_bias)

        return L4_L2_output  # Defuzzyfied Layer
```

 # Layer 5

```
class SummationLayer(keras.layers.Layer):
    def __init__(self, **kwargs):
        super().__init__(**kwargs)

    def build(self, batch_input_shape):
        self.batch_size = batch_input_shape[0]
        # Be sure to call this at the end
        super(SummationLayer, self).build(batch_input_shape)

    def call(self, input_):
        L5_L2_output = tf.reduce_sum(input_, axis=1)
        L5_L2_output = tf.reshape(L5_L2_output, (-1, 1))
        return L5_L2_output  # output layer
```