DEVELOPMENT OF A BREAST CANCER RISK PREDICTION SYSTEM USING ARTIFICIAL NEURAL NETWORKS (ANNs)

BY

OCHOR EMMANUEL CHERECHI
GOU/U22/CSC/833

DEPARTMENT OF COMPUTER SCIENCE

FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY

GODFREY OKOYE UNIVERSITY, ENUGU, NIGERIA

Dr. Franklin Uchenna Okebanama
Supervisor

MAY 2026

APPROVAL PAGE

This research, titled "**DEVELOPMENT OF A BREAST CANCER RISK PREDICTION SYSTEM USING ARTIFICIAL NEURAL NETWORKS (ANNs)**," has been assessed and approved by the Department of Computer Science Committees in the Faculty of Computing and Information Technology, Godfrey Okoye University, Enugu, Nigeria.

Dr. Frank Okebanama	**.....**	**.....**
Supervisor	Signature	Date
External Examiner	**.....**	**.....**
External Examiner	Signature	Date
Dr. Frank Okebanama	**.....**	**.....**
HOD, Department of Computer Science	Signature	Date
Prof. I. A. Ajah	**.....**	**.....**
Dean, Faculty of Computing And Information Technology.	Signature	Date

CERTIFICATION

I certify that I completed the research included therein and that it was primarily based on my observations and investigation. Consequently, I will fully accept the University's decision if I am found to have cheated on the assignment.

_____

OCHOR EMMANUEL CHERECHI
GOU/U22/CSC/833

DEDICATION

This project is dedicated to all women who have experienced breast cancer symptoms but lacked timely access to proper medical screening, early diagnosis, reliable health information, and adequate care, with the hope that technology-driven solutions will continue to support awareness, early detection, and better health outcomes for women everywhere.

ACKNOWLEDGMENTS

I sincerely acknowledge God Almighty for His grace, strength, wisdom, and guidance throughout the course of this research work.

My profound gratitude goes to my supervisor, who is also the Head of Department, Dr. Frank Okebanama, for his academic guidance, constructive corrections, patience, and encouragement throughout the development of this project. His support and direction contributed greatly to the successful completion of this work.

I also wish to appreciate all the lecturers in the Department of Computer Science and Mathematics for the knowledge, mentorship, and academic support they have given me throughout my period of study. Their teaching and guidance have played an important role in shaping my academic growth and professional development.

My heartfelt appreciation goes to my beloved parents, Prof. Francis Ikechukwu Ochor and Lady Ahunna Cecilia, for their love, sacrifices, prayers, encouragement, and constant support throughout my educational journey. I also appreciate my siblings for their care, motivation, and understanding.

I am equally grateful to my friends and well-wishers, especially Chiemelie Ilodiuba and Uchegbu Promise Ekene, for the support, encouragement, and kind words offered during the course of this project. To everyone who contributed in one way or another to the success of this work, I say thank you and God bless you.

ABSTRACT

Based on data showing that more than 2.3 million new cases were recorded in 2022, Breast cancer has topped the list of the most commonly diagnosed cancers among women internationally. This data implies practically, that smart tools for accessing the risk of breast cancer have become very important, more especially in locations where there is limited or no groundwork for conventionally screening these women. Convention risk models rely on narrow predictor sets and population-specific assumptions, even though they are clinically established. An example is the Gail/BCRAT tool, which cannot capture complex non-linear interactions among clinical variables. Also, many studies in this field focus on comparing algorithms and fail to develop a practical software solution. This research creates a breast cancer classification model based on biomarkers utilizing an artificial neural network (ANN). The input/output platform is a web-based prototype. A shallow multilayer perceptron model with two hidden layers was developed, and it was trained using the Adam optimizer, binary cross-entropy loss, dropout regularization, and the early stopping technique on a publicly available Breast Cancer Coimbra dataset from the UCI Machine Learning Repository. The dataset consists of 116 case-control records and each record is described by nine anthropometric and serum biomarkers. After training, it was evaluated against logistic regression, support vector machine, and random forest baselines. On the held-out test split, the ANN has attained an accuracy of 82.61%, and precision and recall of 84.62%, specificity of 80.00%, and an ROC-AUC of 0.86; the ANN was better than the other three baselines. Still, these figures may have significant variation due to the limited data sample and they should be interpreted as only indicative. The trained model with the processing scaler used was a part of a Flask web application that Beyond user authentication and structured data entry features, also supports the interpretation of results based on probabilities, keeps users' assessment history, and generates reports. The findings show that a shallow ANN can extract significant relationships in small structured biomarker data and it can also be part of a browser-based tool that can be used by people very easily. This tool could be a step in the direction of modelling research and decision support that can be deployed. Still, it is merely a proof-of-concept system which identifies existing cases and controls rather than predicting future risk and should not be equated with a clinical diagnostic instrument; the next steps in research should be focused on larger and prospective datasets, external validation, cross-validated performance estimates, and probability calibration before considering any real-world use.

TABLE OF CONTENTS

LIST OF TABLES

TABLE **PAGE**

LIST OF FIGURES

FIGURE **PAGE**

CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

Breast cancer is the most commonly diagnosed cancer among women worldwide. According to the most recent World Health Organisation estimates, around 2.3 million new cases were recorded in 2022, together with an estimated 670,000 deaths in that year (WHO, 2026). The disease occurs across every region, yet its outcomes are closely bound up with disparities in public awareness, access to screening, timeliness of diagnosis, and continuity of care. The global burden of disease is significant not only with mortality but also for the long-term residual physical, psychological, social, and economic burden of the disease for patients, families, and health systems. This epidemiological burden in turn further strengthens the case for the development of decision-support tools that can facilitate earlier risk stratification plus more personalised prevention pathways. Breast cancer does not have one causative agent. It is established that the development of the disease is linked to the interaction of demographic reproductive hereditary hormonal metabolic and lifestyle factors. Per the guidance given by the CDC and WHO, the key factors relevant to the individual risk are age, female sex, family history and inherited mutations (i.e. BRCA1 and BRCA 2), dense breasts, parous or nulliparous status, hormonal influences, alcohol intake, postmenopausal obesity and inactivity. But, both sources of reference imply that a lot of women who get breast cancer might not have any known risk factors. In fact, it is quite difficult to make a thorough, straightforward decision only based on a few guidelines or rules of thumb (Center for Disease Control and Prevention, 2026). This fact has prompted the development of standardised breast cancer risk models. Traditional risk tools, such as the Breast Cancer Risk Assessment Tool (bcrat) provided by the National Cancer Institute (NCI), estimate a woman's risk of invasive breast cancer within a specific period of time based on epidemiological variables like self-reported reproductive and family history. While they have proven to be clinically valuable to guide counselling and screening recommendations, they have limitations in their utility. The National Cancer Institute (2013) states that the Gail/bcrat model may not be accurate in certain racial/ethnic groups and is not validated in women with prior breast cancer, strong family history, and certain phenotype/genotype profiles. As the population screening database developed, the importance of expanded risk factors, including breast density, hysterectomy, menopausal status, and ethnicity, was seen in the Breast Cancer Surveillance Consortium

(Breast Cancer Surveillance Consortium, 2013). Due to the shortcomings of traditional statistical modelling, new methods of artificial intelligence and machine learning have become a viable option in medicine. In breast cancer databases, machine learning technologies have been utilized for analysing laboratory biochemical values, genetic information, tabular clinical data, and mammographic image data. A recent systematic review has shown that breast cancer prediction models based on machine learning reached a summary area under the curve (AUC) of 0.73. Besides, neural network models showed somewhat stronger pooled performance than non-neural alternatives in included head-to-head comparisons. However, the same literature review also revealed that there still exists heterogeneity, poor calibration reporting is frequent, and risk of bias recurs, meaning that performance alone is not sufficient; methodological quality and implementation practicality also count (Gao *et al.,* 2022). In particular, artificial neural networks are pertinent to this study since they are equipped to learn highly complicated and nonlinear mappings between inputs and outputs. Risk factors for breast cancer rarely add up in a simple manner, e.g. Age could be interacting with menopausal status; metabolic markers could be interacting with body mass index; family history could be changing the importance of other clinical features. That's why, ANN-driven modelling is the method of choice if the goal is to build a predictive model from the clinical data exhibiting multidimensional relationships. Besides, such selection is also justified for undergraduate research from a teaching perspective since it gives an opportunity for the project to exhibit data preprocessing, model designing training validating just-in-time deployment of user interface, all these within a single integrated software product (Hussain *et al.,* 2024; AlSamhori *et al.,* 2024).

1.2 Statement of the Problem

Despite breakthroughs to a lesser degree in scanning and people's knowledge of breast cancer, most breast cancer cases only come to light after significant biological changes have occurred. This is mainly true for situations where regular scanning is not consistent or one's risk level is not determined in an organized manner.

a. Problems with traditional age-based risk evaluation: Simply basing screening on age fails to identify the variety of breast cancer risks. And well-known tools often rely on a limited number of factors and make assumptions about the population that may not perform well in diverse groups. The National Cancer Institute has pointed out the limitations of the Gail/BCRAT model for specific subgroups (National Cancer Institute,

2026), while reviews started to show that many ML studies are themselves high risk of bias, not showing calibration properly, and poorly validated externally (Gao *et al.*, 2022).

b. From predictive models to real systems: Much of the work published in research ends with performance comparison tables and does not produce a functional decision-support tool that a clinician, researcher, or health worker could actually use. Besides, recent implementation studies have raised issues about reproducibility, trust, governance, legal accountability, and integration into the real world. So, the problem is not only to develop an accurate model but also to come up with a simple, transparent, and practical system that shows the logic of ANN-based risk prediction so that it is workable for an undergraduate project and can be developed further as a prototype (Goh *et al.,* 2025; Ghasemi *et al.,* 2024).

c. Neglecting contextual differences in low-income and African countries: Most models today have been created using mainly Western populations, Western imaging facilities, or data ecosystems that do not easily adapt to other contexts. Besides, studies on African women indicate that they have different risk factors which make the use of the foreign models directly problematic (Adeoye *et al., 2022*). This way, there is a real need for methods that not only rely on affordable and easily accessible data inputs but also offer simple implementation without the need for expensive imaging operations (Macaulay *et al.,* 2021).

1.3 Aim and Objectives of the Study

This research aims at developing a breast cancer risk prediction tool using an artificial neural network. Its detailed objectives are to:

i. review existing breast cancer risk prediction tools and machine learning methods to define the gaps which the proposed system will fill;
ii. create an ANN prototype capable of calculating breast cancer risk based on a few clinical and demographic factors;
iii. carry out a training and validation strategy which fits table-style health data;
iv. measure the performance of the ANN with some baseline models to evaluate its real-world usability;

v. design a web-based user interface through which the model can receive inputs and provide a risk estimate for supporting decisions.

1.4 Significance of the Study

If this system does get developed successfully, its value will be felt by many different groups of people who have been briefly mentioned here:

a. Women and other patients who may be at risk: Really, these are the final users who, through this tool, can get support for earlier risk awareness and timely referral, even though a clinical diagnosis is not replaced.
b. Doctors and other healthcare professionals: they benefit from a secondary tool that enhances their skills during the triage, counselling and awareness processes without replacing their expert judgement (Gao et al., 2022).
c. Health institutions that operate in conditions where resources are scarce: they will find a cost-effective, online decision support tool that only demands simple computing and inputs in tabular form which are accessible; no expensive imaging equipment is needed.
d. Officials and policymakers in public health: they come to understand how risk-based stratification of populations might be used to achieve greater efficiency of screening as compared with the universal approach.
e. Students and researchers in future times: they get hold of a definitive, end-to-end gear that touches upon data processing, ANN building, and performance evaluation, as well as launching for real-world machine learning work in health informatics (Efthimiou *et al.,* 2024).
f. The department and the rest of the academic fraternity: they get hold of a local piece of reference that shows how machine learning could be taken up for intelligent clinical systems.
g. Software programmers and innovators in the field of health technology: they are shown an example of how an ANN model can be an integral part of a scalable and readily deployable web application instead of being a stand-alone experiment only (Goh *et al.*, 2025).

1.5 Scope and Limitations of the Study

This study focuses solely on creating a prototype breast cancer risk prediction system by utilizing secondary data and supervised learning techniques. To be precise, it relies on ANN forecasting of structured variables and does not deal with full-fledged diagnostic interpretation of mammographic images. As per the project's instructions, the research paper focuses more on system designing, model development, and testing, with no hypothesis frame or clinical trial validation involved. That means, as indicated by Collins *et al.* (2024) and further endorsed by Moons *et al.* (2025), the output here will be a proof-of-concept decision-support system rather than a medically certified device. And the study is based on the availability of data. Very limited are the publicly available breast cancer risk model data having truly long horizons of a prospective nature. Most open datasets focus on diagnosis rather than risk prediction, or they are small retrospective case-control samples (Gao *et al.,* 2022). That means the dataset chosen might be estimating "risk prediction" as a probability from health indicators rather than a full-fledged epidemiological risk in the strictest sense (Hussain *et al.,* 2024).

1.6 Operational Definition of Terms

a. Breast cancer risk prediction: The term in this research means using computer methods to calculate the odds of someone being breast-cancer-positive or in the higher-risk group based on a few chosen variables instead of a final clinical diagnosis.
b. Artificial neural network (ANN): This refers to a kind of machine learning model under the guidance of a trainer that consists of computational units connected like neurons of a brain, which learn the weighted relations between input features and an output class or probability.
c. Supervised learning: This means the process of training the model using the data examples where the output, i.e., the right label for each example, is already known.
d. Feature engineering: This term implies gathering the variables needed for the prediction with respect to transformation, selection, or construction so that the model can learn better.
e. Sensitivity: This refers to the ratio of true positive cases identified by the system correctly, and specificity is the ratio of true negative cases identified correctly.
f. AUC: This is short for the Area Under the receiver operating Characteristic curve and measures how well a model can distinguish between classes at different threshold settings (Alba *et al.,* 2017).

1.7 Organisation of the Study

This research study is divided into five main sections. The first chapter introduces the topic by providing background information, outlining the problem to be addressed, stating the main aim and specific objectives, highlighting the importance, delimiting the study, defining operational terms, and giving the layout of the entire research. Chapter Two makes a literature review of conceptual, theoretical, and empirical works related to breast cancer risk AI, ML, and ANNs. Chapter Three outlines the methods used to develop the proposed prediction system. It is concerned with the choice of data source preprocessing model design, training procedures, evaluation strategy, and system development activities. The penultimate chapter (four) discusses how the new system was implemented and presents the results obtained, while the final chapter (five) synthesizes the entire project work by drawing reasonable conclusions and giving recommendations for future research in the field of ANN and medical predictive applications.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter brings an overview of the existing literature in the area of breast cancer risk prediction using artificial neural networks. Initially, the review discusses conceptual issues related to breast cancer, risk factors, risk modelling, artificial intelligence, and neural networks. Later, it explores the theoretical basis of ANN-based prediction, reviews the ANN concept for risk modelling, and integrates findings of experimental studies on tabular, biomarker, and imaging-based datasets. At the end, the chapter summarizes the main issues that motivate the current research.

Breast cancer risk modelling is a very interdisciplinary research area, involving population health, screening art, and computational pattern identification. At a different stage, the field has come up with statistical tools that consider only a few variables to multimodal AI systems that not only integrate clinical information and images but also exploit longitudinal data (Gao *et al.*, 2022). The relevance of this transformation to our study is two-fold: For one thing, it reveals the huge potential of ANN-based approaches and however it highlights the demand for architectures that are explainable and feasible in simple project settings (Mendes *et al.,* 2025).

2.2 Conceptual Review

2.2.1 Overview of Breast Cancer

An uncontrolled proliferation of abnormal cells in breast tissue leads to breast cancer. These cells not only form tumours but can also invade surrounding tissues or metastasize to distant organs. WHO notes that breast cancer often originates in the lining of the milk ducts or lobules and that early stage or in situ cancer can be treated more effectively than invasive cancer. A person's chance of survival hinges largely on the detection of the disease at an early stage followed by accessible treatment. Because of this, the notion of risk prediction gains significance tremendously if an individual's high risk can be identified early, the likelihood for them to be monitored, to change their lifestyle, or to be referred to a specialist becomes greater (WHO, 2026).

2.2.2 Breast Cancer Risk Factors

It is a common practice to classify breast cancer risk factors as either non-modifiable or modifiable (WHO, 2026). Non-modifiable factors entail a person's age gender having relatives with the disease, carrying mutated genes, having very dense breasts, prior breast problems, as well as a few reproductive events. Modifiable factors correspond to drinking alcohol, getting obese after menopause, being inactive, taking some types of hormones, and certain features of reproductive behaviour (CDC, 2026). Modern reviews highlight that breast cancer risk comes to be changed and diversified as biological, environmental, and behavioural factors interact at various stages of the life cycle (Nicolis *et al.,* 2024). A good understanding of this level of intricacy helps in arguing for models that, unlike the fairly simple scoring schemes based on addition of effects only, are capable of representing multi-factor, non-linear interactions.

2.2.3 Breast Cancer Risk Prediction

Risk prediction is about determining the likelihood of a person developing breast cancer within a specific time frame or of them being in a high-risk category. In hospitals or clinics, the goal is not to achieve 100% accuracy but to be able to classify patients. Take the NCI's Gail model, for instance; it calculates a woman's risk for breast cancer over the next five years and over her lifetime, based on her medical, reproductive, and first-degree family history. The BCSC has gone a step further in this direction by adding such factors as breast density and biopsy history to the breast cancer risk assessment through mammograms on a very large scale. These are very useful tools but at the same time they reveal a really big problem in the field. On one side, practical models must be simple enough to be usable, but, they must be sufficiently rich to capture real biological complexity (National Cancer Institute, 2026).

2.2.4 Artificial Intelligence in Medical Decision Support

Artificial intelligence (AI) in medical decision support means computer systems that help doctors or patients by first picking up on patterns, then coming up with forecasts, and finally showing suggestions based on data. When talking about breast cancer treatment, AI has been employed not only for mammography screening and identification of suspicious lesions but also to assist with reading pathology slides, modelling cancer recurrence, and predicting risk in the future (AlSamhori *et al.,* 2024). Most of the time, authors of reviews mention its ability

to enhance personalisation and effectiveness of the workflow but, Then again, advise that one should not be so quick to adopt as model bias and inability to generalise well to other settings, as well as difficulty in understanding the model's decisions, often means a decrease in trust and can also aggravate the problem of inequality (Goh *et al.*, 2025).

2.2.5 Machine Learning in Disease Prediction

Machine learning is an excellent choice for disease prediction since it is able to learn complex or large datasets without the need for the modeller to indicate all the interactions beforehand. In breast cancer, machine learning has been used with patient demographic information, routine laboratory parameters, features extracted from images, and last but not least genetic data (Gao *et al.,* 2022). However, aggregate evidence from meta-analyses indicates that improved average ability to discriminate between cases and controls does not necessarily result in a better clinical utility (Liu *et al.,* 2025). The characteristics that determine a model's usefulness include: good calibration, validation by independent samples, fairness in subgroups, and open reporting.

2.2.6 Artificial Neural Networks

Artificial neural networks (ANNs) are one of the types of machine learning models inspired from biological nervous systems and in training phase, they build constant updates of weights of connections between neurons as a network. Breast cancer risk researchers are keen on ANNs because they can discover complex non-linear association between predictors that fine linear approaches could miss in analysis (Gao *et al.,* 2022). Per a literature review article, neural networks-based models for breast cancer risk prediction generally demonstrated stronger pooled performance than other machine learning algorithms in some contexts, mostly those where imaging 9 data was integrated, although problems such as overfitting and unexplainability still remained (Ghasemi *et al.,* 2024).

2.3 Theoretical and Conceptual Frameworks

2.3.1 Pattern Recognition Theory

Pattern recognition theory serves as the main theoretical foundation for this study. A prediction model receives feature measurements, identifies recurring patterns among these features, and

uses them to tell categories apart. For example, in breast cancer several features like age, body fat, hormone exposure, family history, blood marker, and image texture can together form a clinically relevant pattern. The works of literature based on AI keep referring to pattern recognition capabilities as one of the most powerful attributes of breast imaging and risk assessment (AlSamhori *et al.,* 2024), In particular in circumstances where subtle signs are nearly impossible to manually combine (Ahn *et al.,* 2023).

2.3.2 Supervised Learning Framework

The other system is supervised learning, in which a system model receives labelled instances and learns to associate input parameters with an output result. A set of predictor attributes are given as input, and the breast-cancer class is the target output that the model is trained to predict. This conceptual setup also has a methodological aspect since it influences the way data is split, loss is minimized, and unseen case generalization is performed (Collins *et al.,* 2024; Efthimiou *et al.,* 2024).

2.3.3 Clinical Risk Modelling Framework

Another learning system is supervised learning in which a system model is given labelled instances and learns to connect input parameters with the output result. A set of predictor attributes is provided as input, and the breast-cancer class is the target output that the model is trained to predict. This conceptual structure has also a methodological dimension because it affects how data is divided, loss is reduced, and generalization to new cases is done (Collins *et al.,* 2024; Efthimiou *et al.,* 2024).

2.3.4 Conceptual Framework for the Proposed Study

The conceptual system of this research is illustrated by a four-stage process. Firstly, selected variables connected to breast cancer are extracted from a secondary dataset. Secondly, these variables are pre-processed through cleaning scaling encoding, and feature selection. Thirdly, the processed features are fed into an ANN that maps inputs to the binary output of risk. Fourthly, the final model is integrated into a user interface that receives input data and delivers a predicted risk class or probability. Two quality-control loops (validation and usability) are

placed at the periphery of this flow. Validation is used for checking model performance (Gao *et al.,* 2022), while usability is aimed at making sure the output is understandable enough so that it can be used as decision support rather than as an obscure automation (Goh *et al.,* 2025).

2.4 Artificial Neural Network Fundamentals for Risk Prediction

The main parts of an artificial neural network (ANN) are the input layer, one or more hidden layers, and the output layer. Neurons function by first receiving input values, then multiplying the inputs by respective weights that they've learned, summing up the results, adding bias, applying an activation function, and finally passing the results forward. When a network is training, it checks its predicted outputs against the actual labels, calculates the loss, and changes the weights with backpropagation. This is a good neural network set-up for breast cancer prediction because it enables different variable combinations to jointly exert effects instead of being obliged to have strictly independent relationships.

Weights and biases are key because they decide how much each feature, after learning, will contribute to output. To put it simply, the network does not get informed beforehand which variables are the most important. On the contrary, it discovers these patterns by training on the data. This is part of why artificial neural networks are a good choice for health data where interactions are not always clear (Gao *et al.,* 2022), but also part of the reason why they are much less interpretable than the coefficients in logistic regression (Ghasemi *et al.,* 2024).

Non-linear learning is made possible by the use of activation functions. It is common to use ReLU in the hidden layers as an activation function because it is very simple and supports efficient gradient-based first-order optimization (Nair & Hinton, 2010), while a sigmoid function is naturally suited to the binary output layer since its output is a value in [0,1] that can be interpreted as a probability. As far as optimization algorithms go, Adam has become the first choice for many researchers since it makes use of adaptive moment estimations and, at the same time, is computationally efficient (Kingma & Ba, 2015). Given that medical data sets tend to be small, it is Mostly important to use regularization techniques like dropout (Srivastava *et al.*, 2014) or early stopping that prevent overfitting by limiting the ability of the network to memorize training noise. These methods are available from standard deep learning libraries.

ANNs provide great benefits for breast cancer risk prediction on their ability to model complex nonlinear relationships in a flexible way and to take different kinds of input data. At the same

time, their weaknesses, as outlined by Gao *et al.* (2022), Ghasemi *et al.* (2024), and Goh *et al.* (2025), include the need for extensive data, the dependence on the quality of preprocessing, the lack of transparency, and the risk of overfitting with small data sets. Even though ANN-based and deep learning models frequently work well, there are still issues for explainability, subgroup fairness, and deployment that have not been solved, based on the authors.

2.5 Empirical Review of Related Studies

The empirical evidence related to this research crosses clinical data, biomarker research, mammographic imaging, longitudinal modelling, and system deployment. The individual pieces of work are analysed based on their topical content below and involve table 2.1 for consolidation.

Works relying on clinical, demographic, or laboratory data establish the pros and cons of ANN-type architecture models. Patrcio *et al.* (2018) utilized patient's age, body mass index glucose *resistin* and similar biomarkers to detect breast cancer, proving that routine blood parameters could be a strong support for prediction modelling. In addition, their study created the Coimbra dataset, which this paper uses. Sepandi *et al.* (2018) built an ANN to support radiologists in the estimation of risk and presented their study results for sensitivity at 0. 82 and specificity at 0. 90, which is quite good. However, Salod *et al.* (2019) discovered random forest did better among a number of competitors on the Coimbra dataset by which ANN can not be superior all the time, at least in the tabular setting 74% accuracy, for data. This point of benchmarking is also raised in the wider comparative study: Islam *et al.* (2020) made a comparison support vector machine, k-nearest neighbours, random forest, ANN, and logistic regression on the Wisconsin dataset and kept the top three classifiers, while Rabiei *et al.* (2022) inspected random forest, a multilayer perceptron, gradient boosting, and genetic algorithms in demographic, laboratory, and mammographic features. These papers together raise the point of the appropriateness of the ANN model being benchmarked to the logistic regression, support vector machine, and random forest baselines rather than being reported in isolation.

The biomarker data modelling through neural models is a separate line of inquiry in this field. Through employing principal component analysis with a feedforward neural network on the Coimbra dataset, Yavuz and Eyupoglu (2020) achieved an average accuracy of 97.73%, which is extraordinarily high for such a small clinical dataset. At the end of 2023 Essa Ismaiel, and Al Hinnawi (2024) utilized a convolutional neural network with features derived by hand

for the Coimbra and Wisconsin datasets and still got very high accuracy figures. Still, despite the positive aspect of these findings, the highly accurate results obtained from very small data sets lead us to a cautious note that such performance may be influenced by the method of data splitting and That's why over-estimated internal performance Mainly when validation is not stringent enough (Gao *et al.,* 2022).

Another aspect of research relevance and interpretability is the third issue (Macaulay et al. 2021): Machine learning-based breast cancer risk prediction for African women was designed by them as a bespoke model to illustrate how risk profiles in different African populations might be different enough to warrant a local adaptation; while the focus was not on the use of ANN, it was only incidental to the context that they highlighted the lack of models being transplanted across populations without taking into consideration cultural, reproductive, and epidemiological differences. As a follow-up to this, Islam et al. (2024) made SHAP analysis to explain the workings of an explainable machine learning model, giving rise to the viewpoint of prediction being made clear, trustworthy, and not opaque. By these two investigations, the present paper's focus on implementation, local adaptability, clear circumscription, and easy-to-understand results finds direct support.

Imaging studies have continued to advance this area by showing that mammograms offer risk indicators that are not limited to the usual breast-density measurement. Yala *et al.* (2019) demonstrated that a new hybrid deep learning model that incorporated full-field mammograms with traditional risk factors reached an AUC of 0.70, a level that is better than Tyrer-Cuzick at 0.62 and a random-forest/logistic-regression benchmark at 0.67. It was a great change, because it proved that information on risk-related factors can be obtained from images directly. The work by McKinney *et al.* (2020) was a further development implementation at the large-scale screening level by releasing an artificial intelligence system that performed comparably to or even better than radiologist readers in both UK and US datasets. The follow up step was external and multi-institutional validation: Yala *et al.* (2022) released a report showing that the Mirai model retained its performance when tested with diverse sets from seven hospitals across five countries, while Omoleye *et al.* (2023) studied Mirai in a population with high ethnic and racial diversity which was enriched for African American women and high-risk cases. In their study, 1-year and 5-year AUCs were 0.71 and 0.65, respectively. These works point the research beyond mere claims of performance at a single site to the much tougher proposition of generalization.

Recent articles point towards hybrid solutions as the way forward. Combining mammographic assessment through CNN with clinical variables extracted from EHRs, Michel *et al.* (2023) achieved an AUC of 0.654 which in general, did not represent a significant increase over the BCSC model but, Then again, it was remarkably better in certain racial and ethnic groups. Avendano *et al.* (2024) tested Mirai on a group of Mexican women and observed the performance was moderate overall, with the average C-index and AUC of about 0.63 while the number of cancers in women who were above the assigned high risk level was almost three times that in the lower risk group. Tendero *et al.* (2026) carried out studies of hybrid AI strategy on the cohort of Spain's screening program and concluded that the combined model had the strongest 2-year AUC at 0.75, just slightly better than the image-only and the structured-data models. Put together, these investigations strongly suggest that breast cancer risk modelling's future will be multimodal while their results show that performance is still highly dependent on the population and data.

Another significant direction is longitudinal modelling. Melek *et al.* (2023) and Wang *et al.* (2025) looked at spatiotemporal or multi-time-point mammography models. They argued that the variations that occur with each repeated screening examination can be used to further improve both short- and long-term risk prediction compared to single-time-point models. Although this research is of great scientific importance, it is quite expensive on resources and it is very much reliant on longitudinal imaging archives. For an undergraduate software project using only structured data, it is more appropriate to treat this as an indicator of the future direction of the field rather than as a direct methodological example.

System-oriented breast cancer applications are still relatively underdeveloped compared to other areas, though there has been some effort recently to address this imbalance. Dada *et al.* (2024) described a web application for breast cancer prediction powered by support vector machines, whereas Lin *et al.* (2024) created a web-based AI clinical decision support system for predicting breast cancer recurrence. Such works demonstrate that converting predictive modelling into interactive systems is possible but they also indicate that most applications are still centred on diagnosis or recurrence while only a few focus on risk prediction using ANN.

Table 2.1: Literature Summary Table

Study	Title	Method / data	Key contribution
Patrício *et al.* (2018)	Using resistin, glucose, age and BMI to predict the presence of breast cancer.	Serum biomarkers (Coimbra)	Blood variables support prediction
Sepandi *et al*. (2018)	Assessing breast cancer risk with an artificial neural network	ANN; clinical risk factors	Early ANN risk estimation feasible
Salod *et al.* (2019)	Comparison of the performance of machine learning algorithms in breast cancer prediction and diagnosis	Multiple ML (Coimbra)	Random forest best; ANN not assured
Yala *et al.* (2019)	A deep learning mammography-based model for improved breast cancer risk prediction	Deep learning; mammograms	Risk signal beyond breast density
Islam *et al.* (2020)	Breast cancer prediction: A comparative study using machine learning techniques	SVM/KNN/RF/ANN/LR (Wisconsin)	Supervised classifier comparison
Yavuz & Eyupoglu (2020)	An effective approach for breast cancer diagnosis based on routine blood analysis features	PCA + feedforward NN (Coimbra)	High accuracy on small data
McKinney *et al.* (2020)	International evaluation of an AI system for breast cancer screening	Deep learning; UK/US screening	Matched or exceeded radiologists
Macaulay *et al.* (2021)	Breast cancer risk prediction in African women using random forest classifier	Random forest; African women	Population-specific modelling needed

Rabiei *et al.* (2022)	Prediction of breast cancer using machine learning approaches	RF/MLP/GBT/GA; clinical data	Neural and tree models effective
Michel *et al.* (2023)	Breast cancer risk prediction combining a convolutional neural network-based mammographic evaluation with clinical factors.	CNN + EHR hybrid	Helps under-served subgroups
Omoleye *et al.* (2023)	External evaluation of a mammography-based deep learning model for predicting breast cancer in an ethnically diverse population	Mirai validation; diverse cohort	External validation essential
Avendano *et al.* (2024)	Validation of the Mirai model for predicting breast cancer risk in Mexican women	Mirai validation; Mexican women	Performance drops across populations
Essa *et al.* (2024)	Feature-based detection of breast cancer using convolutional neural network and feature engineering	CNN + features (Coimbra)	Engineered features aid classification
Islam *et al.* (2024)	Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI	ML + SHAP (explainable AI)	Transparent, interpretable predictions
Tendero *et al.* (2026)	Breast cancer risk assessment for screening: A hybrid artificial intelligence approach	Hybrid AI; clinical + imaging	Best short-term discrimination

These studies really show how three things happen again and again. One, neural and deep learning methods are not always better; how well they seem to work is very much up to things like the kind of data, how much data, how the data is checked, and who the data is about (Gao

et al., 2022). Two, what some people say is the difference is very confusing: some works predict who will get the disease, some identify who already have the disease or cancer, and identification changes completely the reasoning of this study and also partly explains why public data sets and student works are often at the level of probability-of-present systems when real prospective risk datasets are hardly accessible. Three, only a very small number of studies actually make their models into usable software (Hussain *et al.,* 2024). The current study takes on this last, long-standing issue by incorporating a well-tested ANN into a user-friendly and open web-based tool.

2.6 Gaps in the Literature

2.6.1 Methodological Gaps

The first major methodological gap is the most critical one. To illustrate, thorough systematic reviews have revealed that most of the machine learning breast cancer risk prediction studies tend to only present discrimination findings and generally ignore calibration, in-depth bias evaluation, or strong external validation (Gao *et al.,* 2022). This is quite a big flaw since risk prediction models have to not only be good at separating classes, but also provide probabilities that are sufficiently meaningful to support actual decision-making (Alba *et al.,* 2017). In fact, poor calibration has been highlighted as one of the key issues in the recent reviews (Moons *et al.,* 2025).

2.6.2 Data and Contextual Gaps

The other gap is about data and context. What is more, almost all of the top recent mammography studies have been done in mammography-rich environments and mostly Western or multi-institutional datasets Still datasets which are open and contextually relevant to low-resource settings are very rare. Given that a study on women in Africa has already shown that simply transferring models from other populations may be inappropriate (Macaulay *et al.*, 2021) So a gap for practical work, i. e. studies that make use of accessible health indicators and openly available data yet are clear about the limits of generalisation remains (Avendano *et al.,* 2024).

2.6.3 Accessibility and Usability Gaps

The last gap identified in the literature relates to usability. Predictive models may perform well, but few studies move beyond algorithmic evaluation to the production of a working interface for end users. Besides, the implementation research of late has also shown that there are still issues about reproducibility trust governance, and explainability (Ghasemi *et al.,* 2024). In the absence of clear and comprehensible outputs, it becomes quite challenging to establish the real-world utility of a model, and breast cancer is in fact one of the scenarios where both users and doctors strongly prefer AI to be a helping tool rather than a decision-maker (Goh *et al.,* 2025).

2.6.4 Contribution of the Present Study

By presenting an Artificial Neural Network (ANN)-based breast cancer risk prediction model that can feasibly be carried out within the limitations of an undergraduate computer science course, this paper touches a lot on some of the gaps mentioned. It includes the major part of the work with structured data, clearly explained preprocessing, evaluation with no sole reliance on raw accuracy, and a simple web interface that can be quickly deployed. Even so, it does not claim to solve the issue of external validation across all populations nor to replace clinical tools that are well-established. Instead, it is simply a hands-on development study transforming the ANN-based breast cancer prediction into a complete system design with limitations clearly stated.

CHAPTER THREE

METHODOLOGY AND SYSTEM DESIGN

3.1 Introduction

This chapter covers the methodological structure, data processing methods, model development, system architecture, and evaluation setup of the breast cancer risk figure prediction system. The theme of the chapter is centred on the development of a software product rather than on running tests of hypotheses. So, the focus is on what manner of analysing design implementation training integration, and evaluation of the system will be both defendable technically and achievable practically. Following the guidance of the model prediction development, the chapter starts with the context and requirements then moves to data handling, model specification, validation, and reporting considerations (Collins *et al.*, 2024).

3.2 System Analysis

What has been done in research tissues using breast cancer risk support tools point to the potential of prediction but also point to the shortcomings of the present practice. To illustrate, a classic example is the calculator of the National Cancer Institute that determines invasive breast cancer risk from personal and family history factors like age, reproductive history, number of first-degree relatives with breast cancer, previous biopsies, and atypical hyperplasia. The tool is clinically effective, but it is based on a limited set of predictors and it should still be used with caution in particular for some special subgroups and clinical histories (National Cancer Institute, 2026). Besides, systematic review of machine-learning-based breast cancer risk models discloses that although a large number of studies demonstrate good performance, these reports still lack calibration of heterogeneity, poor reporting of calibration, and limited clinical utility (Gao *et al.,* 2022).

3.2.1 Analysis of the Existing System

To put it simply, the "existing system" for the present investigation amounts to three rather imperfect alternatives: informal clinical judgement, conventional statistical calculators, and

stand-alone machine-learning studies published in journals. When critically viewed, their features show these weaknesses:

a. **Informal clinical judgement**: depends on the individual's knowledge and insight, so it is very difficult to standardize, it can be almost impossible for it to be reproduced by different individuals, and assessments may differ drastically.
b. **Conventional statistical calculators (e. g. Gail/BCRAT):** these can be used by many people as they are accessible, yet they are constrained to using only the fixed statistical assumptions, they only work with a limited set of predictors, and they are not able to model complex non-linear interactions among clinical variables.
c. **Population-specific design of conventional tools**: majority of the conventional tools were calibrated for a particular population, that is why their estimates may not be reliable when applied to different demographic or clinical settings.
d. **Stand-alone machine-learning studies**: in general, these studies limit themselves to publishing algorithmic performance results instead of proceeding to build an application with features like structured data entry, prediction, and record storage.
e. **Weak translation into practice**: both informal judgment, calculators, and machine learning approaches are still unable to be implemented due to barriers like lack of trust, different generalizability, and integration into real clinical workflows.

3.2.2 Architecture of the existing System

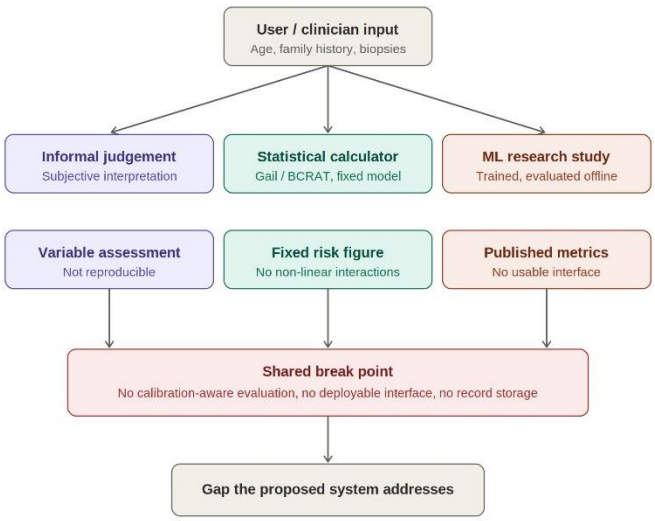

Figure 3.1: Architecture of the existing system

Figure 3.1 presents the system architecture of the currently available approach of breast cancer risk assessment. It is not a single cohesive system, but rather three independent pathways sharing the same initial inputs: informal clinical judgement, fixed statistical calculators such as Gail/BCRAT, and stand-alone machine-learning studies. Each of them only produces one type of output (an unstandardized clinical judgement, a fixed risk number, or a set of published performance metrics) without any of them being transformed into a practical tool. All three culminate at the same point of failure: the absence of calibration-aware evaluation, a transparent deployable interface, and record storage. The proposed method precisely aims at this shared lack.

3.2.3 Limitations of the Existing System

a. Limited in modelling non-linear relationships: a lot of the current tools do not model non-linear complex interactions of metabolites and demographic factors very well.
b. Population dependence: multiple popular tools were actually created for certain populations and may be hard to adapt to new populations.
c. Poor deployment: a significant percentage of the breast cancer AI research has not been turned into simple, user-friendly software.
d. Explainability and confidence gaps: making the outcomes explainable and increasing the trust of the predictions are still the key challenges in AI-supported breast cancer diagnosis.

3.3 System Design

Development-oriented research design is the way used in the study, where the primary product is a working prototype for structured-data risk estimation. This approach is quite sensible considering that the project is focused on transforming expert knowledge and a secondary dataset into an operational prediction system which gets its inputs through an artificial neural network and produces an interpretable output. Given this premise, the design is both constructive and reflective: it is about making a product and slowly improving it; the main way the effectiveness is determined is through technical assessment rather than social-science hypothesis method.

3.3.1 System Design: Agile Methodology

Agile comes with short cycles, quick models, nonstop improvement, a lean mentality of software over paperwork, etc. The Agile Alliance's values put emphasis on interactions between people, functional software, collaboration with customers, and responding to change, while Scrum illustrates the way complex products can move forward step by step with lots of planning, reviewing and re-scheduling (Agile Alliance, 2001). In this particular case, Agile is kept as a very practical tool for work planning rather than a formal company practice: one sprint for requirements and wireframes, one for data preparation, one for ANN modelling, and a last one for interface integration and testing.

Besides, web-based systems are even more exposed to the benefits of Agile due to the continuous need for UI adjustments, constant feedback from users and most importantly, the ability to support staged deployment, which means that features can be tested and continuously improved throughout development.

3.3.2 ANN Architecture and Design Rationale

The ANN to be implemented will be a small multi-layer perceptron that is best suited for tabular data. In the input layer, there will be one neuron per predictor that has been kept after the preprocessing stage. The hidden structure will, at the beginning, be made up of two dense hidden layers, while during the tuning, different candidates like 16-8, 32-16, or 64-32 neurons might be tested. Since the Coimbra dataset is small, a simpler architecture will be employed; using very deep networks would mainly lead to overfitting while there may not be much increase in generalization.

ReLU activation will be used in the hidden layers as it is not only computationally the simplest, but is also an effective function for gradient-based learning. However, the output layer will feature sigmoid activation which outputs the probability of the positive class in the range 0 to 1. Since this is about binary classification, the loss function is binary cross-entropy and for the optimizer, Adam is the choice as it is efficient and has been shown through experience that it performs well across various deep learning tasks. Dropout in the hidden layers, L2 penalty, and early stopping based on validation loss will form the regularization. The design is purposely very simple: the aim is to get stable results on small tabular data rather than architectural novelty.

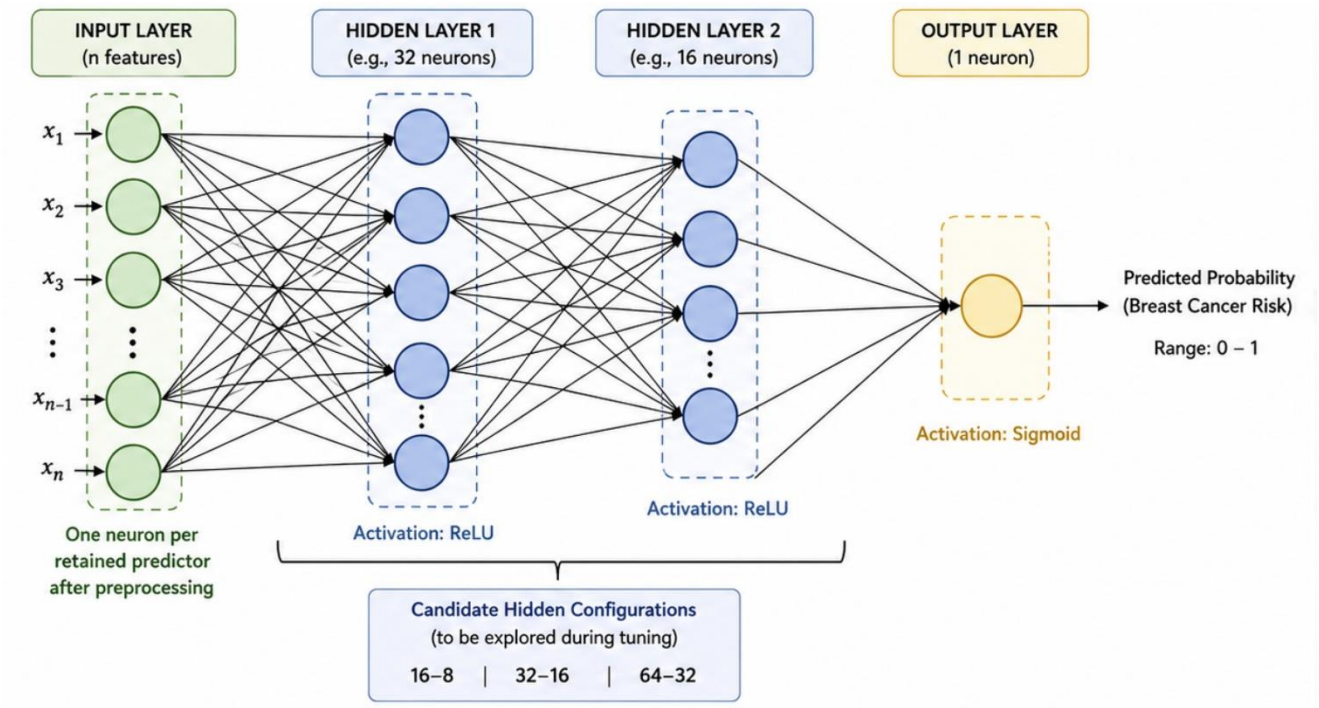

Figure 3.2: Proposed ANN Architecture for Breast Cancer Risk Prediction

3.3.3 System Requirements Specification

Functional Requirements include the system: identifying users who have access through login, offering users a way to choose different predictor variables, verifying the data completeness and the correctness of the data range, performing the preprocessing steps, executing the trained Artificial Neural Network (ANN) model, and ultimately generating a probability score and risk class as output. Also, the users should be able to interact with the system through a web browser, enter information by filling in forms on the web pages, and get the results change on the screen dynamically as the backend works through their requests. The system will Also keep a record of the assessments and enable the generation of reports ready for printing.

Non-Functional Requirements entail the system being easily understandable with minimal training needed for navigation, having a web interface that reacts very quickly to user requests, requiring less computational power, and being able to maintain very short inference times. It must be online at all times so users can have consistent access; fairly secure via input validation and simple authentication; make consistent sense internally; and be clear and straightforward in its phrasing. Since decision-support tools should not be the only basis for care decisions, the output page will show a short cautionary note explaining that the result is supportive and not

definitive, which is a principle that matches the current cancer risk tools and the main AI-related issues that have been raised in breast care (National Cancer Institute, 2026).

3.3.4 Data Sources and Data Collection

The secondary data that will be used to develop the system is from the Breast Cancer Coimbra dataset, which is available at the UCI Machine Learning Repository. The dataset includes information of 116 women, of which 64 are breast cancer patients and 52 are healthy controls, and it was given in 2018. It is multivariate, health-related, and has been explicitly made available for classification work by using anthropometric variables and routine blood-analysis parameters (UCI Machine Learning Repository, 2018). Based on the study by Patricio et al. (2018) which was the original basis of the dataset, combinations of age BMI glucose, and *resistin* were found to be potentially good low-cost predictors of breast cancer presence.

3.3.4.1 Source of Data

The source is a publicly available tabular dataset and not an imaging archive. That is a deliberate choice. Recent studies indicate that image-based models can outperform non-image models on average; But suitable imaging datasets that are openly accessible for risk modelling are very difficult to use in a small undergraduate project, and on top of that, many breast cancer risk prediction machine-learning studies still suffer from issues in approach despite their greater performance claims (Gao et al. 2022).

3.3.4.2 Data Acquisition Procedure

The data will be collected in CSV format by downloading it from the repository. It will then be imported into a Python environment and arranged in a tabular form for further handling and modelling. Since the repo has not announced any missing data, the data acquisition step tends to be simple; Even so, reproducible scripts will be responsible for checking file integrity, data types, variable names, and class labels before the training phase is undertaken (UCI Machine Learning Repository, 2018).

3.3.4.3 Inclusion and Exclusion Criteria

Exclusion Criteria are those records that have been found to be duplicates, corrupted entries, impossible numeric values upon validation, or any record whose target label cannot be mapped consistently to the modelling pipeline. In fact, only a few exclusions are expected as the source dataset is small and has already been curated (*Oza et al.,* 2020).

3.3.4.4 Data Description

It is difficult to overstress the value of the Coimbra dataset for an undergraduate ANN project. It is publicly available, well-structured, and small enough to allow the students to carry out full-cycle experimentation. Though, its limitations are just as obvious: the number of samples is small, and the task should be considered as prediction from accessible biomarkers rather than a large-scale prospective population risk model. Such a limitation will be explicitly recognized in the interpretation of system outputs (UCI Machine Learning Repository, 2018).

3.3.4.5 Description of Input Variables and Output Variable

The input variables are age, body mass index glucose insulin, HOMA *leptin adiponectin, resistin,* and MCP-1. They are a mixture of demographic, metabolic, and inflammatory indicators that have been linked to breast cancer prediction in previous studies. The output variable is a binary class denoting the presence or absence of breast cancer in the dataset (Patrcio *et al.,* 2018).

3.3.5 Data Dictionary

Table 3.1: Data Dictionary Table

Variable	Description	Type	Proposed role
Age	Age in years	Numerical	Demographic predictor
BMI	Body mass index	Numerical	Anthropometric predictor
Glucose	Blood glucose level	Numerical	Metabolic predictor

Insulin	Blood insulin level	Numerical	Metabolic predictor
HOMA	Homeostasis model assessment	Numerical	Insulin resistance indicator
Leptin	Serum leptin level	Numerical	Adipokine / biomarker
Adiponectin	Serum adiponectin level	Numerical	Adipokine / biomarker
Resistin	Serum *resistin* level	Numerical	Inflammatory biomarker
MCP-1	Monocyte chemoattractant protein-1	Numerical	Inflammatory biomarker
Classification	Breast cancer class label	Binary	Output variable

The choice to use these variables is in agreement with the UCI dataset documentation as well as the original research paper of Coimbra that pointed out age BMI glucose, *resistin*, and other elements as major predictors of breast cancer (UCI Machine Learning Repository, 2018).

3.3.6 Data Preprocessing and Feature Engineering

Data cleaning will initially involve recognizing data types, identifying duplicates, and computing basic statistics. Given the dataset's limited size, every record counts and This way, only clearly incorrect entries will be heavily scrutinized and possibly discarded. Dealing with missing data is not expected to be a major issue since the UCI documentation states there are no missing values but a checking program will be run to verify this. In case of any unexpected missing values, the project will resort to simple and well-documented imputation methods rather than aggressive ones (UCI Machine Learning Repository, 2018).

Determining how to handle outliers will be a result of using exploratory graphics and checking the data distributions. Removing outliers will not be a default action because some of these values may be extremely valid physiological measures rather than data-entry mistakes. It will but be evaluated whether for the mainly skewed biomarkers, robust scaling, winsorisation, or log transformation is the most appropriate action. Since all selected variables are numerical, encoding for categoricals is not a major step in this dataset. Feature scaling still seems to be a must for an ANN, and so, z-score standardisation will be used to counter the issue of one

variable dominating the optimisation process and to ensure stable optimisation (Kingma & Ba, 2015).

To mitigate noise and overfitting, feature selection will be implemented. The study will enlist the knowledge of the subject matter in the first step, utilizing all variables, then proceed with correlation analysis, variance check, and features importance or ablation tests. Feature engineering like the addition of interaction terms and the use of transformed versions of the skewed biomarkers might be considered if They enhance model stability; But, at the same time, excessive complexity has to be excluded because small tabular health datasets are extremely susceptible to overfitting (Yavuz & Eyupoglu, 2020). As for class imbalance, the dataset is only slightly skewed, so the main way to handle this will be through stratified splitting. Should the effect of imbalance on learning become pronounced, then class weights will be used instead of a heavy synthetic oversampling so that the original data distribution will not be too changed (Yavuz & Eyupoglu, 2020). For Feature Selection, the goal of the project is to keep it simple. With only nine variables, feature selection would mostly be used as the tool to check if performance can be maintained or improved by using less. This stage could involve correlation checking, domain knowledge, and straightforward model-based importance assessment within cross-validation only. The relevant points are sustained by the original Coimbra paper which showed that a small subset of features can perform just as well, if not better, and also by the general principle that prediction modelling should strive for both simplicity and good prediction (Efthimiou *et al.,* 2024).

3.3.7 Training Procedure and Hyperparameter Tuning

Since the dataset is small, the training phase should keep the maximum amount of information possible but at the same time prevent the problem of being overly optimistic.

The best option would be to split the whole dataset into the development set and the final test set in a stratified way 80:20. For hyperparameter tuning and final model selection, a stratified 5-fold cross-validation will be performed on the development set. The method is a nice middle ground for practicability and methodological rigor and is, in fact, even more justifiable than just one random split (Efthimiou *et al.,* 2024). On the training, the pipeline will be: take the training split and do preprocessing; train the hearable ANN; record the validation loss; perform early stopping but load the best weights; and finally repeat with the other splits. The hyperparameters that could be changed are these: number of hidden layers, number of

neurons/layers, dropout rate, learning rate, batch size, number of epochs, and L2 regularisation strength. The search strategy might be grid-based or random search given the computational resources Yet the search space will be small on purpose so that the tuning stage does not become a source of instability. The final model will be the one which has the strongest average cross-validated performance and acceptable calibration on the development data, after which it will be evaluated once on the untouched test set (Alba *et al.,* 2017).

3.3.8 System Modelling Using UML

The standard UML is a formal graphical language to visualize specify build and document software artefacts. Besides, it is ideal to present the proposed system in a clear undergraduate project format. In the current work, UML is used to represent the system from three different perspectives: the user's interaction, the process flow, and the static class structure (Object Management Group, 2017).

3.3.9 Use Case Diagram

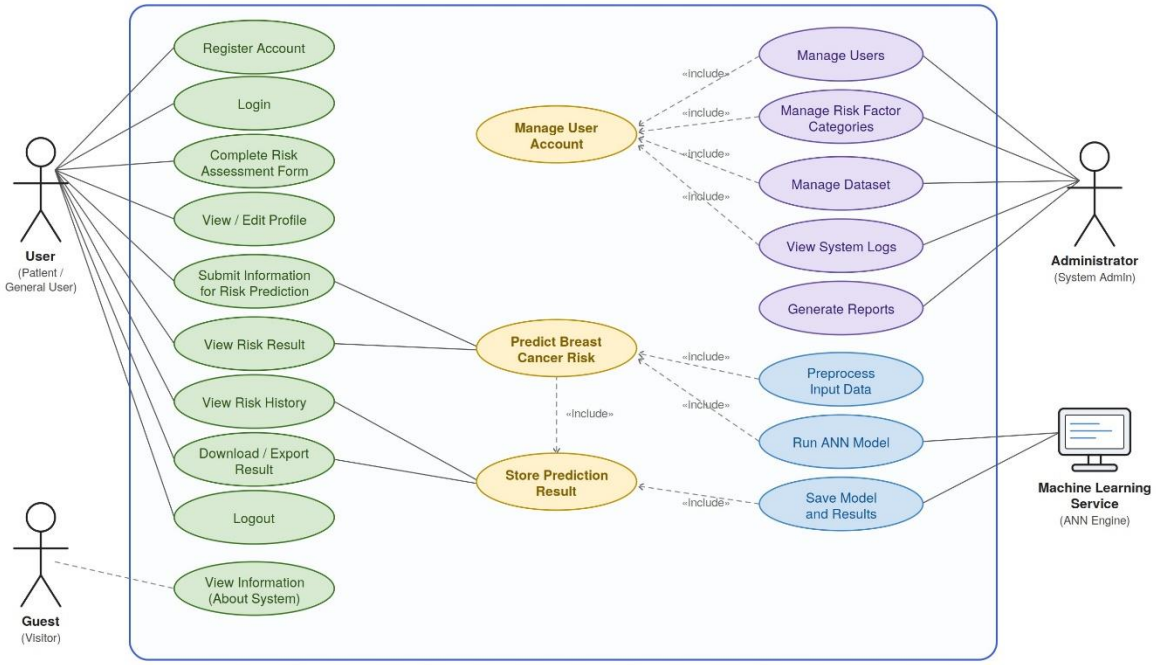

Figure 3.3: Proposed System Use Case Diagram

The use case diagram shows the functional scope of the breast cancer risk prediction system and the actors interacting with it. Four actors are depicted. Three of them are humans: the User (patient or general user), the Guest (unauthenticated visitor), and the Administrator, whereas the Machine Learning Service (ANN Engine) is a non-human system actor. This last one is the trained model and the preprocessing pipeline that together act as an automated participant in the prediction process.

The User initiates the main sequence of events that includes registering, logging in, filling out the risk assessment form, sending the information for prediction, and finally viewing downloading, and revisiting their results. But the Guest is only allowed to see the general information about the system (this is illustrated with a dashed association to indicate the lighter, read-only relationship). The Administrator is the one who controls the core parts of the system: users, risk-factor categories, the dataset, system logs, and reports.

The include relationships reveal the necessary minor steps that operate as part of the major use case. When the system does Predict Breast Cancer Risk, it must include Preprocess Input Data and Run ANN Model (actions performed by the Machine Learning Service), and prediction also includes Store Prediction Result, which in turn includes Save Model and Results. But the administrative side management tasks are modelled as includes of Manage User Account. The use of includes here keeps the common steps defined only once and reused rather than duplicated across every parent use case.

Overall, the diagram clearly separates three concerns: what end users do (the green cluster), what system does internally (the amber core), and the automated machine-learning work delegated to the ANN engine (the blue cluster).

3.3.10 Activity Diagram

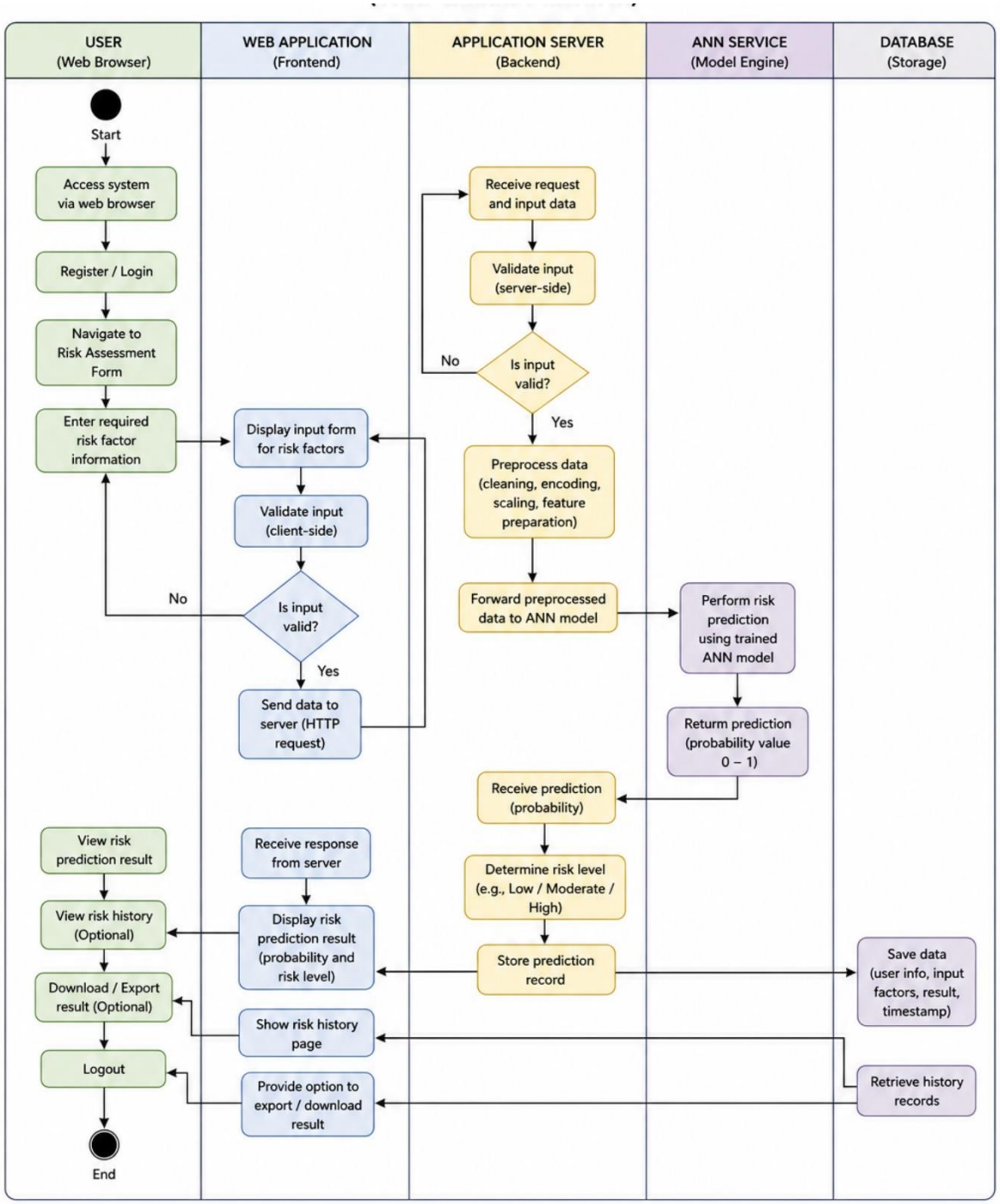

Figure 3.4: Proposed Activity Diagram

This activity diagram illustrates the complete processing of a risk assessment by one user through five swimlanes (User, Web Application, Application Server, ANN Service, and Database) from login and form submission, through two validation checks (client-side, then server-side, both going back to the user on invalid input), and then through the steps of preprocessing, ANN prediction, risk-level interpretation, and storage. It works as a counterpart

to the use case diagram. Instead of showing the capabilities of the system, it graphically depicts how the responsibility is passed from one component to another at runtime. It finishes by showing that the user is able to view, export if desired, and review the stored result before logging out.

3.3.11 Class Diagram

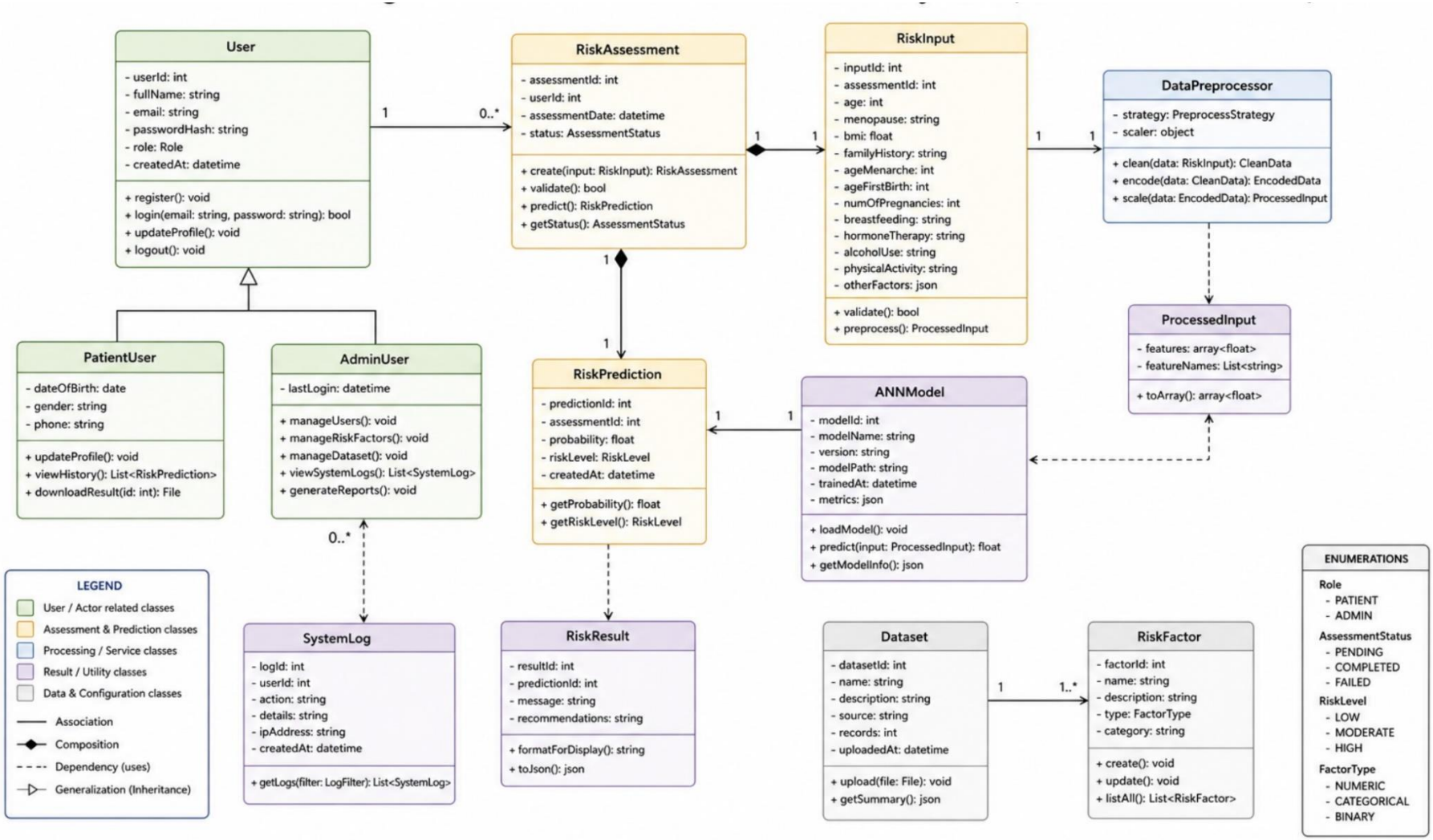

Figure 3.5: UML Class Diagram

The class diagram visually represents the static architecture of the breast cancer risk prediction system including the classes, their properties, operations, and the former ways of establishing relationships for mutual work. Several classes are categorised and colour-coded per their functionalities: user/actor classes are green, assessment and prediction classes are amber, processing/service classes are blue, result and utility classes are purple, and data/configuration classes are grey. The focus of the design is the User superclass that comprises common identity and authentication features (userId, email, passwordHash, role) and user-related operations like register(), login(), and logout(). Through generalisation, two radically different classes inherit from it: PatientUser, which is identified by demographic attributes and can handle results with features like viewHistory() and downloadResult(); and AdminUser, holding various administrative tasks like manageUsers(), manageDataset(), and generateReports(). Subclassing these behaviours keeps the shared behaviour in one place, while they each can extend it as per their role.

The prediction in the workflow is represented by the amber and blue classes and the composition relationships between them. A User can have related to zero or more RiskAssessment objects (1 to 0..*), each RiskAssessment is composed of, without a doubt, one RiskInput element (which holds the collected risk factors such as age, bmi, familyHistory, and reproductive history) and one RiskPrediction. The filled diamond is used to denote composition, signifying that those parts cannot exist independently of the assessment that owns them. DataPreprocessor, is tasked with the operations clean(), encode(), and scale(), generating a processed input, upon which RiskInput is dependent; the latter is then used as input for the ANNModel, whose predict() function delivers the eventual RiskPrediction with probability. Dashed arrows represent these dependency relationships, showing the difference between temporary collaboration and stronger structural ties. Besides methods, the classes also exhibit behaviour and provide utility. RiskPrediction is tied to a RiskResult, which is in charge of producing the output in both display and exporting formats, whereas SystemLog registers all the administrative actions (AdminUser is the creator of 0..* log entries). Also, the grey Dataset and RiskFactor classes symbolize the training data and its adjustable risk factors, where one dataset corresponds to one or more risk factors (1 to 1..*). The enumeration part on the right side is the controlled value sets that are employed throughout the model: Role (PATIENT, ADMIN), AssessmentStatus (PENDING COMPLETED FAILED), RiskLevel (LOW MODERATE HIGH), and FactorType (NUMERIC CATEGORICAL BINARY), thereby making only these valid, pre-set values allowed. In sum, the illustration reveals the manner in

which the system separates user management, data preparation, model inference, and result handling into well-integrated class modules having clearly defined responsibilities.

3.3.12 Input Design and Output Design

Input design will focus on creating a clear and comprehensive form that contains the nine predictor fields. Each field will have a label that describes it, a small note to explain the unit, and validation messages to catch mistakes before data entry is finished. Output design will include showing the predicted probability, the human-understandable risk category, the date and time when the assessment was done, and a short message saying that the system is the user's friend rather than a doctor. This line is very important since a trusted breast cancer AI needs to give data that is useful without resulting in baseless certainty (Goh *et al.,* 2025).

3.3.13 Database Design

The design will be a straightforward relational model with only a few tables: users, patient data, assessments, and reports.

Table 3.2: Database User Table

Field Name	Data Type	Description	Key
user_id	INT	Unique identifier for each user	Primary Key
full_name	VARCHAR(100)	Name of the user	
email	VARCHAR(100)	User email address	Unique
password_hash	VARCHAR(255)	Encrypted user password	
role	VARCHAR(20)	User role (e.g., Admin, User)	
created_at	DATETIME	Account creation timestamp	

Table 3.3: Patient Record Table

Field Name	Data Type	Description	Key
record_id	INT	Unique patient record identifier	Primary Key
user_id	INT	Reference to user who created record	Foreign Key
age	INT	Patient age	
bmi	FLOAT	Body Mass Index	
glucose	FLOAT	Glucose level	
insulin	FLOAT	Insulin level	
family_history	VARCHAR(50)	Family history of cancer	
other_factors	TEXT	Additional clinical inputs	
created_at	DATETIME	Record creation timestamp	

Table 3.4: Assessment Results Table

Field Name	Data Type	Description	Key
result_id	INT	Unique assessment result ID	Primary Key
record_id	INT	Linked patient record	Foreign Key
probability_score	FLOAT	Predicted probability (0–1)	
risk_level	VARCHAR(20)	Risk categorization into: low, moderate or high	
prediction_date	DATETIME	Date that the prediction took place	

Table 3.5: Reports Table

Field Name	Data Type	Description	Key
report_id	INT	Unique report identifier	Primary Key
result_id	INT	Linked assessment result	Foreign Key
report_content	TEXT	Generated report details	
generated_at	DATETIME	Report generation timestamp	

3.3.14 Evaluation Strategy

The evaluation plan is designed to check not only the accuracy of a model's classification but also how much its output probabilities can be trusted. Such a capability, in fact, is a required feature of a risk-support prototype. In this context, the validation approach will be internally cross-validation combined with the final test evaluation on independently held-out data. Methodologically, this alternative to training and testing on the same data is, among other things, recognized as a source of overfitting and too optimistic performance estimates by scikit-learn (scikit-learn developers, 2026d, 2026i).

Performance metrics will include a mix of accuracy precision recall or sensitivity, F1-score, and ROC-AUC for class discrimination, and Brier score and a calibration curve for the quality of measured probabilities. Having such a variety is, in fact, essential because it is perfectly possible for a model to distinguish classes quite well and yet it may not produce well-calibrated probability scores. A recent editorial in The BMJ makes precisely the point about discrimination and calibration difference, and the documentation for the scikit-learn tool shows how to calculate both ROC-AUC and calibration curve for binary classifiers (Riley *et al.,* 2024).

Baseline or Comparative Models will consist of logistic regression and random forest. Logistic regression is a proper baseline since it is a highly interpretable model and has been extensively used in prediction modelling; Then again, random forest is a robust non-linear baseline model for structured tabular data and has already been found relevant to African breast-cancer studies. So, putting the new ANN against such baselines makes more sense than showing the standalone performance of ANN (Macaulay *et al.,* 2021).

In addition, False Positives, threshold-sensitive errors, and cases in the decision boundary will receive the most focus during Error Analysis. Any false negatives produced by the system are going to be investigated for potential concentration of errors to certain biomarker ranges and/or training instabilities due to small sample learning. Also, this type of analysis is entirely in line with the present-day practice of reporting prediction models publicly and, if necessary, subjecting them to scrutiny, rather than boiling them down to just one headline metric (Moons *et al.,* 2025).

CHAPTER FOUR

SYSTEM IMPLEMENTATION

4.0 Introduction

The implementation of a breast cancer risk prediction system is the main point of this chapter. It discloses the software environment tools dataset, neural network architecture, experimental setup, web implementation, testing, and outcome of the entire system. Meeting the aim and objectives setup in the previous chapters, this implementation focuses on the creation of a simple, though functional, decision-support prototype that: can obtain structured clinical and biomarker inputs; process them through an ANN prediction model; deliver a probability based supportive output over the web. The chapter also evaluates the model's performance based on the Breast Cancer Coimbra dataset that was limited and also the prototype which was non-clinical in nature.

Also, it is worth emphasizing that this system should not be considered as a replacement of medical professionals by any means. It is a mere academic prototype illustrating that clinical data and neural networks can be integrated into decision-support workflows. That note is timely since breast cancer AI is a domain where quality explainability transportability, and safety of implementations are still major issues (Moons et al., 2025).

4.1 Development Environment and Tools

Along with machine learning utilities, the software implementation relied on web development tools as well as a lightweight database for support. Python was chosen as the main language mostly because of its huge and diverse ecosystem of libraries for data science, machine learning, and web application development. The setting was designed so as not to burden a normal laptop which was quite a fit given the dataset size and in reality, the work was at the undergraduate level.

Firstly, the implementation stack was composed of four major layers. At the bottom were the data and modelling that handled data loading, cleaning, and model creation plus performance evaluation. Second was the model persistence part saving the trained ANN and the scaler for new predictions. Third came the application logic taking user inputs, validating and

normalizing the data, calling the trained model and formatting the output. Last was the presentation one showing the system via a browser-based interface.

4.1.1 Programming Language and Libraries

Python was selected due to its support for rapid prototyping, simplicity of syntax, re-usability of machine learning pipelines and available libraries for numerical programming and neural network implementation. Pandas and NumPy were used for loading the table of comma separated data, examining variables and managing arrays before transforming the data into numeric feature matrices. Scikit-learn was used for splitting datasets into training and testing, scaling features, developing baselines and scoring models. The artificial neural network was implemented using TensorFlow/Keras since the Keras API makes easier simplicity in constructing feed-forward neural networks, dense layers, dropout regularisation, binary cross-entropy loss, Adam optimiser and early stopping to control the model (Keras 2026a 2026b, 2026d 2026e 2026f, 2026g; scikit-learn developers 2026a 2026b 2026c 2026d, 2026e). For the user-facing application, Flask was used to provide a REST API, and standard web technologies were chosen for the client side.

The pages were built in HTML, with CSS and Bootstrap providing styling for the forms and overall layout. A small amount of JavaScript was incorporated to improve client-side responsiveness. SQLite was used to provide a lightweight relational storage engine for users, prediction history, and reports generated by the prototype. Since this prototype was not intended for use in a substantial clinical system, this technology fit with the design goal of providing a lightweight accessible web-based system.

Table 4.1: Development tools and technologies

Tool/Technology	Role in the System
Python	Core implementation language for data processing, modelling, and backend logic
Pandas	Loading and manipulating the Coimbra dataset
NumPy	Numerical computation and matrix handling

scikit-learn	Data splitting, scaling, baseline modelling, and evaluation metrics
TensorFlow/Keras	ANN design, training, regularisation, and model persistence
Flask	Backend web framework for routing, validation, and inference integration
HTML	Structure of frontend pages
CSS/Bootstrap	Styling layout responsiveness, and presenting forms nicely
JavaScript	Frontend interaction on the lightweight side and validation feedback
MySQL	Storage of user accounts, records of inputs, and history of predictions
Jupyter Notebook	Model development for exploration and analysis by experiment
Visual Studio Code	Developing application code, combining modules, and locating bugs

As a result, these specific options were chosen as a sort of midway point between methodological suitability and practical possibility. Considering the dataset was very small (only 116 strings), it would have been impractical to set up huge infrastructures or specialized GPU-heavy pipelines. A simple ANN and a lightweight web stack turned out to be enough.

4.1.2 Development Platform and Hardware Requirements

The development platform was intentionally kept basic. Since the problem involves clinical and biomarker data in table form rather than high pixel-rate medical images, the model was successfully trained on an everyday laptop-class machine not equipped with a dedicated GPU. This is a huge real-world benefit mostly for educational institutions and small project groups.

Table 4.2: Hardware and software requirements

Component	Minimum Requirement	Reference Development Environment
Processor	Dual-core CPU	Intel Core i5-class processor or equivalent

RAM	4 GB	8 GB
Storage	2 GB free space	10 GB free SSD space
GPU	Not required	Integrated graphics sufficient
Operating System	Windows 10 / Linux	Windows 11
Python Runtime	Python 3.x	Python 3.11
Web Browser	Chrome, Edge, or Firefox	Google Chrome
Development IDE	Any code editor	Visual Studio Code / Jupyter Notebook
Database Engine	SQLite or equivalent	SQLite

The final ANN only had a few trainable parameters, and the dataset itself was very small, so the hardware requirements were hardly raised. Also, from a realization point of view, this is quite the benefit of the prototype: it illustrates that the ANN-based decision making on health variables is possible without expensive-specialized hardware. Of course, such a straightforward development must not be mistaken as clinical readiness. Besides, the actual field deployment would call for more reliable hosting, tighter security, broader validation, more rigorous audit logging, and governance control (Goh *et al.,* 2025).

4.2 Dataset Description and Preprocessing

The Breast Cancer Coimbra dataset downloaded from the UCI Machine Learning Repository was used to develop the model. As was mentioned before, the dataset includes 116 samples, 64 of which are breast cancer patients and 52 are healthy individuals. The explanatory variables considered were Age BMI Glucose, Insulin HOMA *Leptin Adiponectin Resistin*, and MCP-1 whereas output variable was Classification. For this paper, we treated the problem as a binary classification task, so the goal of the model was to predict the probability of a sample being in the breast cancer positive class. This is how "risk" is being used in this work, i.e. it is a measure of probability of belonging to the positive class and so is not indicative of a confirmed clinical diagnosis, nor it is a long-term epidemiological risk value calculated with a formal risk calculator. Initially, the dataset was in a csv file, and then it was loaded to pandas data frame. After that, the variable names were examined, the data types of the nine analysing features were confirmed to be numeric, and the proportions of the two class were also checked. Then

the original target values were altered to binary ones so the data would be easier to model, and in this binary labelling healthy control was set to 0, whereas breast cancer case was set to 1. Using this binary encoding, the data is compatible with the use of sigmoid output and binary cross-entropy loss. The dataset used and the nine variables that were implemented in the prototype are in line with the documentation of the Coimbra dataset and also with the original research by Patricio et al. (2018) which concluded low-cost routine blood and anthropometric variables to be predictive of breast cancer presence. Following the description in the previous chapter, preprocessing was started with an examination of the quality of data. A script-based look was taken to check for missing values, invalid numeric types, and duplicated records. The imported dataset was complete, so there were no missing values. No records were discarded due to incompleteness. A descriptive check revealed that to some extent, biomarker variables, in particular Insulin *Leptin Resistin*, and MCP-1, were more spread out than Age and BMI. Yet, since the dataset is small, removing outliers was not considered a typical preprocessing step. Such a choice was on purpose: in clinical datasets, anomalous values are usually a sign of the real biological variability, and excessive cleaning would simply distort the limited signal available. In fact, the study made use of standardisation and regularized learning to decrease the impact of outliers.

Table 4.3: Dataset variables and preprocessing steps

Variable	Type	Preprocessing Step	Implementation Note
Age	Numerical	Type verification and z-score standardisation	Retained as continuous predictor
BMI	Numerical	Type verification and z-score standardisation	Retained as continuous predictor
Glucose	Numerical	Type verification and z-score standardisation	Retained as continuous predictor
Insulin	Numerical	Type verification and z-score standardisation	Retained despite wide spread
HOMA	Numerical	Type verification and z-score standardisation	Retained as insulin-resistance marker

Leptin	Numerical	Type verification and z-score standardisation	Retained as biomarker variable
Adiponectin	Numerical	Type verification and z-score standardisation	Retained as biomarker variable
Resistin	Numerical	Type verification and z-score standardisation	Retained as inflammatory biomarker
MCP-1	Numerical	Type verification and z-score standardisation	Retained as inflammatory biomarker
Classification	Binary target	Label encoding to 0/1	0 = healthy control, 1 = breast cancer case

All nine predictors were kept in the final training. Such a decision was justified for two reasons. Firstly, the feature space was already very minimal; That's why, there was no need for very radical feature elimination. Secondly, the variables were very closely related to the field and perfectly in line with the main purpose of the dataset. Later, a Standard Scaler was only fitted on the training data and then used for the validation, test as well as the live inference data. The step was very important in preventing data leak because if the scaler was fitted mostly dataset before dividing it, it would have been possible for the information from the cases not seen to affect the training. The standard preprocessing workflow of Scikit-learn is well designed to support this fitting pattern of training-only (scikit-learn developers, 2026a 2026b).

The data were divided through stratified sampling to make sure that the class distribution was kept to be similar across different subsets. The final split that was used for the implementation was: 74 training records, 19 validation records, and 23 records for the final testing. For class, it corresponded to 41 positives and 33 negatives during training, 10 positives and 9 negatives during validation, as well as 13 positives and 10 negatives during testing. There was some class imbalance, but it was quite moderate. So, to prevent learning bias towards the bigger class, we applied balanced class weighting during training. We considered this option to be better than heavy oversampling as the total dataset is very small and synthetic duplication often means overly optimistic results in medical classification studies (Gao *et al.,* 2022).

The last preprocessing issue was related to the deployment. The web application during prediction used the same scaler which was learned during model training, stored, and then

loaded. It made sure that when a user entered values through the web form, those values were transformed in the same exact way as the data that were used to train the ANN. If this consistency had not been kept, the interface would have been separated from the model and the quality of the prediction would have been negatively affected.

4.3 Model Architecture and Training

In this paper, the prediction engine that was built is a feed-forward artificial neural network aimed at tabular binary classification. Considering the very small sample of the Coimbra database, the authors opted for a shallow preview over a deep one. It has already been established that, for very small structured clinical datasets, very large networks tend to memorize the training noise rather than the stable general patterns. So, model simplicity most often prevails as the right solution.

To enable nonlinear learning, ReLU was selected for the hidden layers, and a sigmoid was chosen for the output layer to generate a probability value between zero and one. Adam was used for the optimisation process whilst binary cross-entropy was selected as the loss function. These three are widely documented in the literature of ANN and, in particular, their joint use is one of the cornerstones of the Keras software systems (Keras, 2026b, 2026c, 2026d, 2026e). During the model selection phase, the authors experimented with different hidden-layer setups.

Comparing the effects of hyperparameters on the development data it was found that the 168 architectures offered the best stability-discrimination trade-off. But, increasing the number of neurons to 3216 led to improvement in training accuracy, but the model was more unstable during validation and less tended to generalise well. That means, the 168 setup was finalized as the working model of the paper.

Table 4.4: ANN model architecture

Layer	Output Size / Units	Activation	Regularisation / Note
Input Layer	9	–	One node for each predictor
Dense Hidden Layer 1	16	ReLU	L2 regularisation applied

Dropout Layer	16	–	Dropout rate = 0.25
Dense Hidden Layer 2	8	ReLU	Compact hidden representation
Output Layer	1	Sigmoid	Predicts probability of positive class

The network finally consisted of 305 trainable parameters, which is really tiny compared to the current deep learning standards. Such small size was deliberate and perfectly aligned with the method. The aim was not to invent a new architecture but to create a simple and clear ANN pipeline that can be trained reproducibly on a small health dataset and incorporated into a web application.

Training was done by using a maximum of 100 epochs with early stopping. After each epoch, validation loss was checked, and the best weights were restored once validation performance stopped getting better. Early stopping is crucial in small datasets because it decreases the likelihood that the model will keep learning noise after the useful signal is gone (Srivastava *et al.,* 2014). For this project, the best validation result came at epoch 26 while training ended at epoch 34 when patience was used up. This means that the network reached the point of convergence quite fast and the extended training was not required.

Table 4.5: Training configuration

Parameter	**Setting**
Input Features	9
Hidden Layers	2
Hidden Units	16 and 8
Hidden Activation	ReLU
Output Activation	Sigmoid
Loss Function	Binary cross-entropy
Optimiser	Adam
Learning Rate	0.001
Batch Size	8

Maximum Epochs	100
Early Stopping	Enabled
Early Stopping Patience	8 epochs
Best Epoch Restored	26
Training Stop Epoch	34
Regularisation	L2 + Dropout (0.25)
Class Weighting	Balanced
Classification Threshold	0.50

While the premises of this research work advocated using stratified 5-fold cross-validation with 5 separate random 5-folds for decision making on model selection for the development stage, the ultimate system implementation was a hybrid of this earlier tuning strategy and a static training-validation division for monitoring the learning process. Put simply, 5-fold cross-validation was the first step in hyperparameter optimization. Once the best ANN was determined, the 74-record training subset of samples was used for training, 19-record validation subset of samples was used for monitoring, while the model was finally evaluated on a single occasion using the untouched 23-record test subset of samples. This is in line with the advice from recent prediction-modelling literature that stresses the significance of separating the steps of model selection from the final hold-out evaluation (Efthimiou *et al.,* 2024). The training and validation curves have indicated a reasonable convergence behavior. The training loss was consistently reduced from about 0.69 at the start to 0. 388 at the restored best epoch. Validation loss also followed the same decreasing pattern initially and then reaching a plateau at 0.472.

Training accuracy increased much faster and higher than the validation accuracy, which is very typical with small datasets, but the difference between the two accuracies did not increase so much so that it shows a worrying degree of overfitting. Mostly, the curves were indicative of mild overfitting that was under control rather than overfitting so severe leading to memorisation.

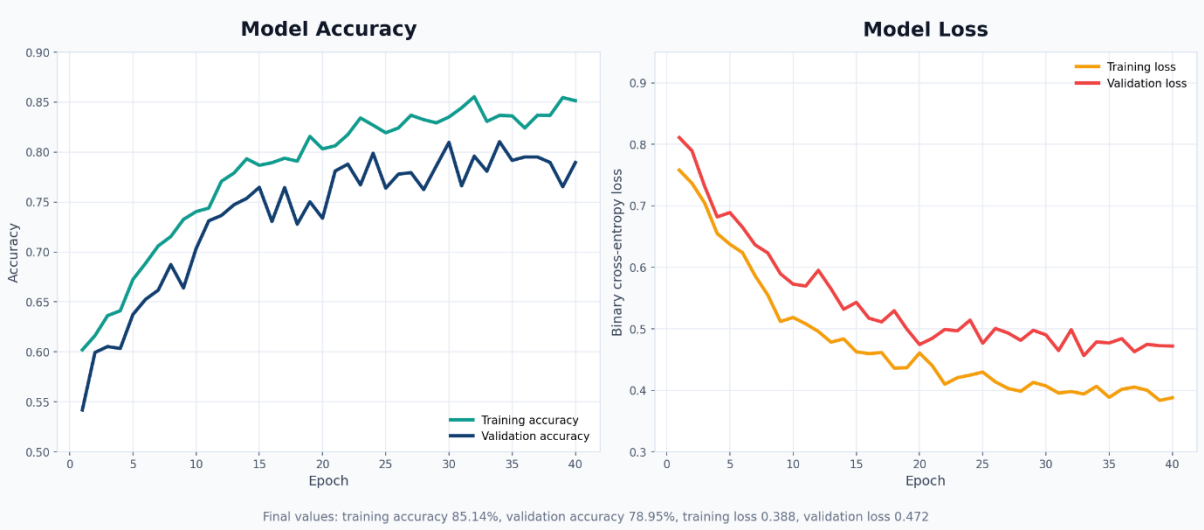

Figure 4.1: ANN Training History for Breast Cancer Risk Prediction

4.4 System Implementation and Modules

After training the ANN model, it and the fitted scaler were saved and imported into the web-based application. This stage of integration basically changed the model from being just a notebook prototype to an interactive web application. The web app layout was set up as per the traditional browser-server architecture. User's input on the client side was acquired using form elements. On the server side, data first went through error checking and were afterward fed to the saved pipeline. Eventually, the ANN model was invoked, the output probability was decoded, the record was saved, and the results page was generated and sent back. Breaking down the system, technologically speaking, into separate functional modules:

1. The first one was the user authentication and management module. The tasks of user registration logins session management, and logouts were managed by this module. To ensure that the risk assessment functionalities were only available to legitimate users, we implemented basic password hashing and authentication checks. This was crucial as the prototype also kept a log of the assessment submissions and report generations. 2. The next module was the risk factor entry module. The module displayed a web form consisting of nine predictor fields that the ANN required: Age BMI Glucose, Insulin HOMA Leptin Adiponectin Resistin, and MCP-1. Each of these variables was clearly labelled and set up to only accept numeric values.

2. Validation cues were added on the form so that any blank entries, letters, or realistically impossible values could be flagged either before or during submission. While the database layout allows for extra optional metadata at a later date, only the nine chosen predictors were used as input to the trained model during prediction.

3. The last module was responsible for preprocessing and making predictions with the model. At this stage, the form entries were first changed into a numerical vector, sorted per the order of the features, and then transformed using the stored *StandardScaler* right before being handed over to the ANN. The model output was a probability score ranging from 0 to 1. For classification, a cut-off value of 0.50 was adopted. Any score equal or greater than the threshold was regarded as an indication of a positive class, whereas those below the threshold were regarded as negative. And this binary decision, the system also provided a probability statement to support the output so that it would not seem like a strict yes/no decision.

4. Result interpretation module is the fourth module. This is the module that interprets the raw model output and transforms them into a format understandable to non-technical users. The results page showed the probability score, the predicted class, the time of prediction, and a warning advisory note. Besides that, the interface also divided the score into user advisory bands for better understanding. Scores lower than 0.40 corresponded to a low-risk prediction, scores between 0.40 and 0.69 indicated a medium-risk prediction, and scores from 0.70 and above represented a high-risk prediction. This communication layer was separate from the binary decision threshold. Its purpose was solely to help users understand. And, the output page clearly stated that the result is not a medical diagnosis and should not replace clinical judgement.

5. The history, reporting, and administration module formed the fifth component. Once a prediction was made and its inputs validated, the probability score, interpretation, and timestamp and the input variables were saved in a SQLite database. Users could retrieve their past assessments through the history page. Also, report create tool was kept simple so that the users could easily print risk summary evaluations. Having admin privileges also meant that the administrators could trace user actions and results audit logs, which are useful for monitoring the prototype and making presentations.

In software engineering terms, the development of the modules was done bit by bit. The first sprint had us bringing up a prediction form and getting the off-line model ready. The main model was the focus of the next sprint during which the model was made available via Flask routes while the scaler and model files were serialized. Almost everything, including better feedback messages, database persistence, and navigation among different pages, was accomplished during the final sprint. This iterative process matched Agile orientation, which was mentioned in the beginning as a research approach, and it turned out to be very helpful as both the artificial neural network and other elements of the interface still had to be kept on the go during development.

4.5 Testing and Evaluation

Tests were conducted along 2 broad lines. One, the ability of the artificial neural network (ANN) to predict and the relative performance of other basic models. Two, the software correctness, the interface usability, and the integration quality. This twofold evaluation method was essential since the project's purpose was not only to train a model but also to create a functional prototype system.

4.5.1 Evaluation Metrics

Evaluating the model was done with a range of common binary classification metrics derived from confusion matrices. Firstly, accuracy was used to calculate what fraction of predictions were right overall and was defined as adding up true positives and true negatives, then dividing by the total number of cases (observations). Precision was the ratio of correct positive predictions to all positive predictions made. Then again, recall or sensitivity was the fraction of positive cases that the model correctly identified. Specificity was the fraction of negative cases correctly identified by the model.

The F1-score was the harmonic average of precision and recall and was very helpful in cases when both metrics needed to be in the balance. The ROC-AUC metric was the overall score blind to the actual threshold between how well the model ranked the positives higher than negatives. These are the mathematical representations of the indicators:

In formula form, the main indicators are:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision = TP / (TP + FP)

Recall or Sensitivity = TP / (TP + FN)

Specificity = TN / (TN + FP)

F1-score = 2 × (Precision × Recall) / (Precision + Recall)

Within this research, the positive class is the breast-cancer-positive class of the Coimbra dataset. Apart from the discrimination metrics above, the Brier score was also used to measure the quality of probability together with the other indicators. A model may have a good discrimination power but the probabilities it outputs may be quite poorly calibrated (scikit-learn developers, 2026f, 2026g).

4.5.2 System Testing

Functional and integration tests were conducted to analyse if the software worked correctly from entering the data to delivering results. Black-box testing was performed on the major modules like registering, logging in, validating forms predicting saving history, and generating the report. Besides that, browser-based integration tests were executed which were testing that the ANN prediction path was always in line with the archived preprocessing pipeline. This was a pretty big deal because if the system used one way of normalization at the time of training and a different one at the time of the actual prediction, the output would be totally wrong even if the web system would appear to be working.

Table 4.9: System module testing

Module	Test Case	Expected Outcome	Observed Outcome	Status
Registration	Submit valid new user details	Account should be created successfully	Account created and redirected to login page	Pass

Login	Submit valid credentials	Dashboard should open	User authenticated and dashboard displayed	Pass
Login validation	Submit wrong password	Access should be denied	Error message displayed and login blocked	Pass
Prediction form validation	Leave required fields blank or enter text in numeric field	Submission should be blocked	Inline and server-side validation messages displayed	Pass
Preprocessing pipeline	Submit nine valid inputs	Stored scaler should transform inputs correctly	Input vector processed without dimensional or type errors	Pass
ANN inference	Submit complete validated record	Probability and class should be returned	Prediction returned successfully, average response time 0.84 seconds	Pass
Result storage	Save completed prediction	Record should be written to database	Record stored with timestamp and user ID	Pass
History retrieval	Open prediction history page	Previous records should be listed	Records displayed correctly in reverse chronological order	Pass
Report generation	Generate printable summary	Report should contain inputs, score, result and note	Report generated successfully	Pass
Logout	End active session	Session should terminate securely	User redirected to login page and session cleared	Pass

System testing results indicated that, from a functionality perspective, the prototype had achieved a level of stability appropriate for demonstrating the concept. Running a test on the reference computer, the average time from a valid submission to displaying the result was under one second. This is a very good time of response for a decision-support tool with minimum user interaction. Along the way, minor issues like the proximity of buttons and erroneous timestamp fields in exported reports were identified and fixed before the final evaluation. This kind of iterative fixing process is in line with the refinement approach of Agile.

4.5.3 User Interface Testing and Walkthrough

The testing of the user interface involved a task-based walkthrough of the main screens and workflows. The primary focus was to see whether the prototype communicated its message clearly, permitted the inputs to be made in the right way, gave interpretable outputs, and had navigation kept to the minimum. The walkthrough corresponded to the typical user journey starting with the login and ending with the prediction review.

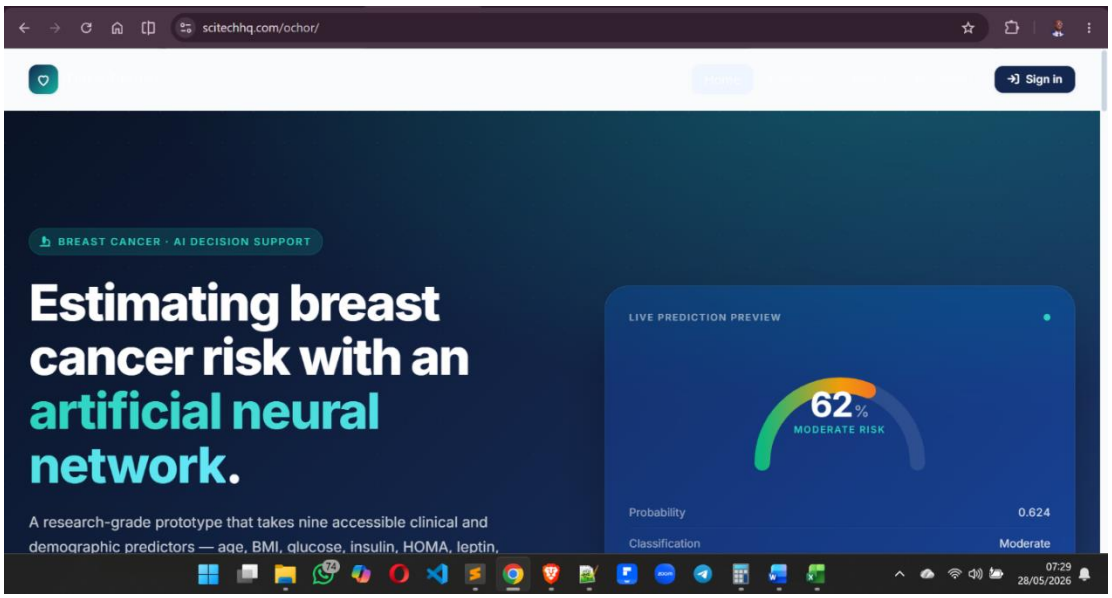

Figure 4.2: Home Page Screenshot

To begin with, login page was tested. It was quite simple and consisted of two main entry fields for e-mail and password, a button for logging in and a link for user registration. Validation messages worked to inform the user when some of the required fields were left empty or invalid credentials were used.

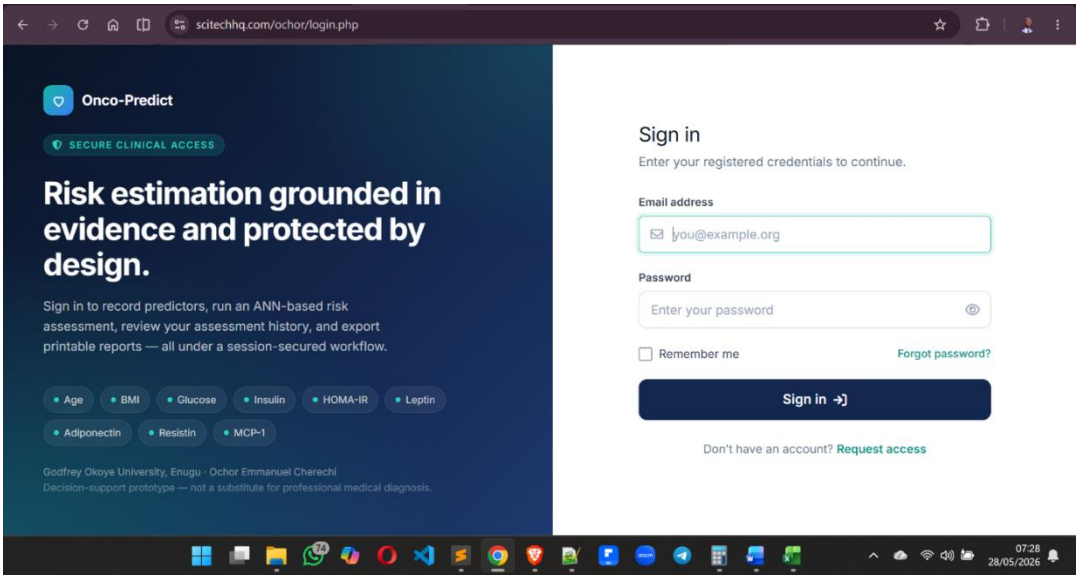

Figure 4.3: Login Page Screenshot

After logging in, the user panel was the very first thing the user saw after a successful login. In fact, the panel served as a starting point to other areas of the system such as the risk assessment form, prediction history, and logout options. The user's primary need, i. e. to carry out a breast cancer risk assessment, was deliberately made so accessible on the dashboard that the user could do it with a minimum of two clicks.

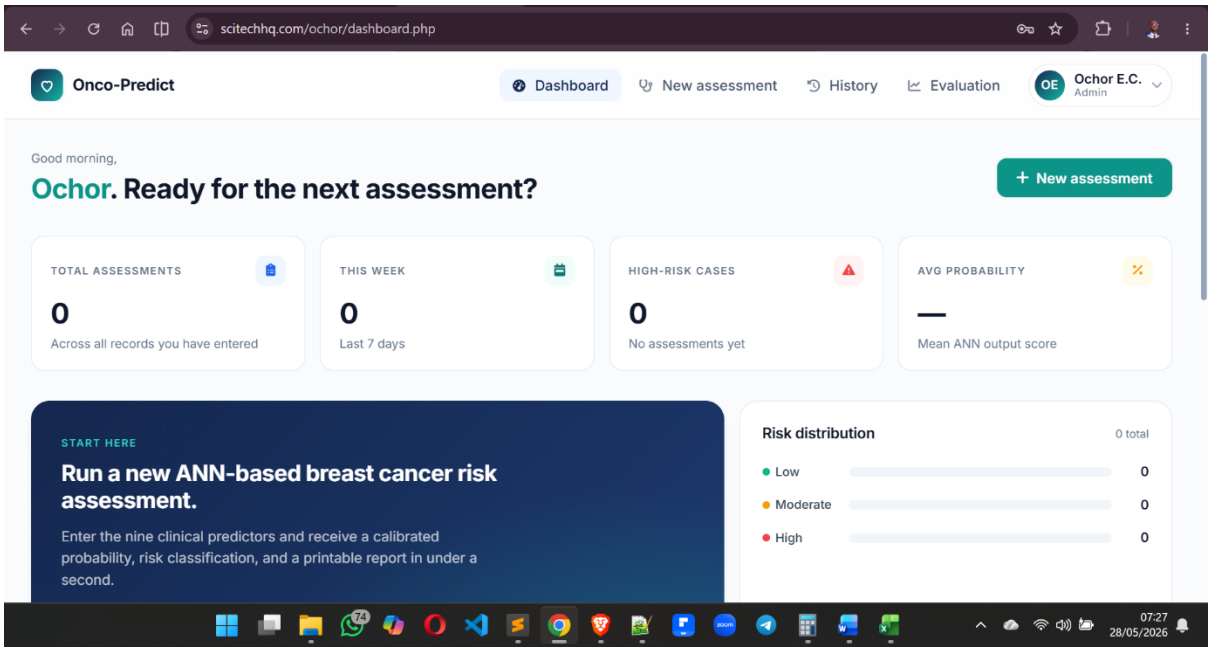

Figure 4.4: Dashboard Page Screenshot

Prediction input form was the next focus of the walkthrough. It contained the entire list of nine variables required by the ANN. Each entry field was clearly marked and set one below the

other to improve readability. Only digits were allowed as input: validation rule. And, upon being pressed, submit button checked model-side validation so no using browser-side validation might mean wrong model input.

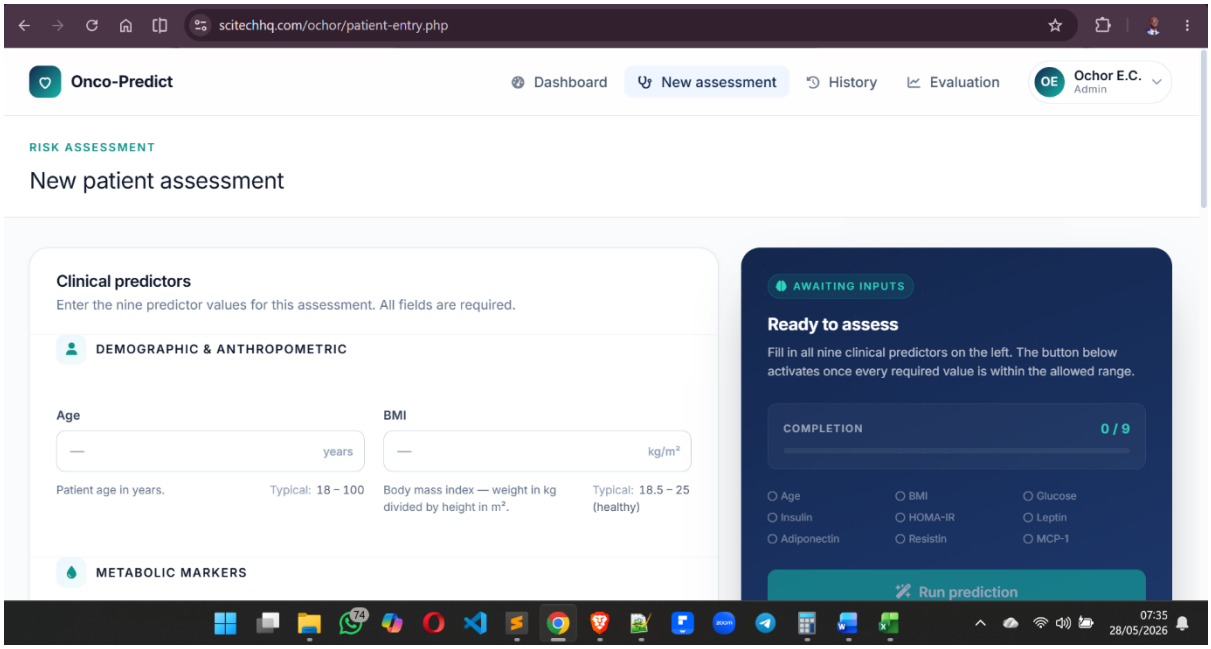

Figure 4.5: Prediction Input Form Screenshot A

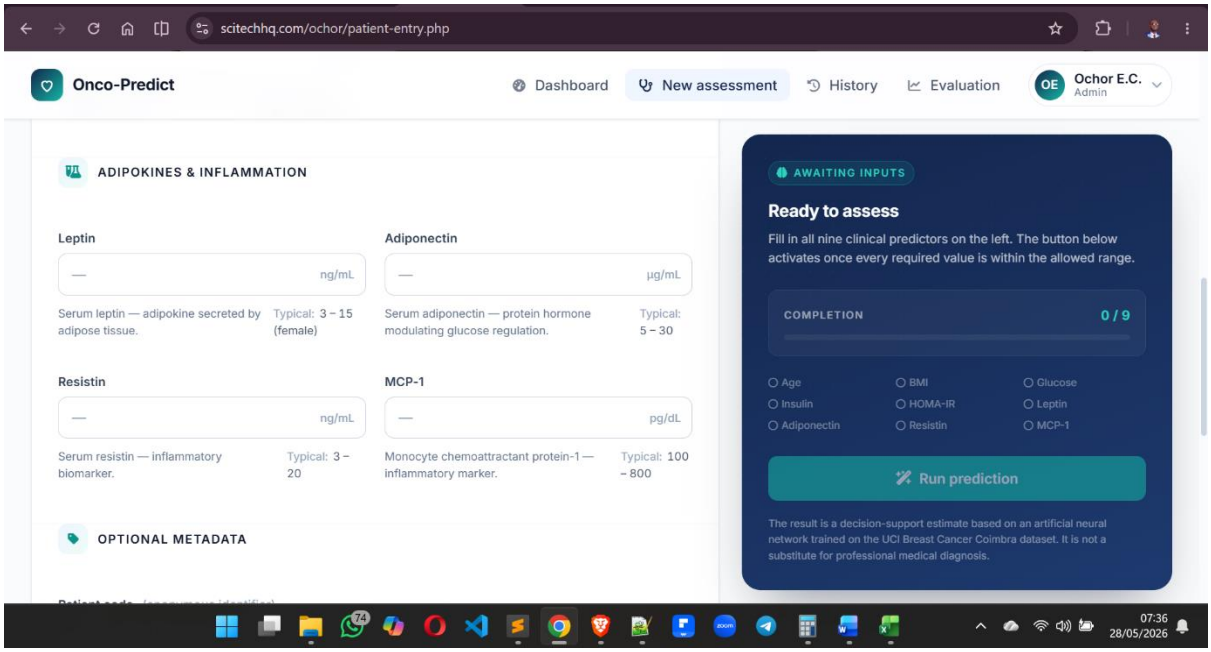

Figure 4.6: Prediction Input Form Screenshot B

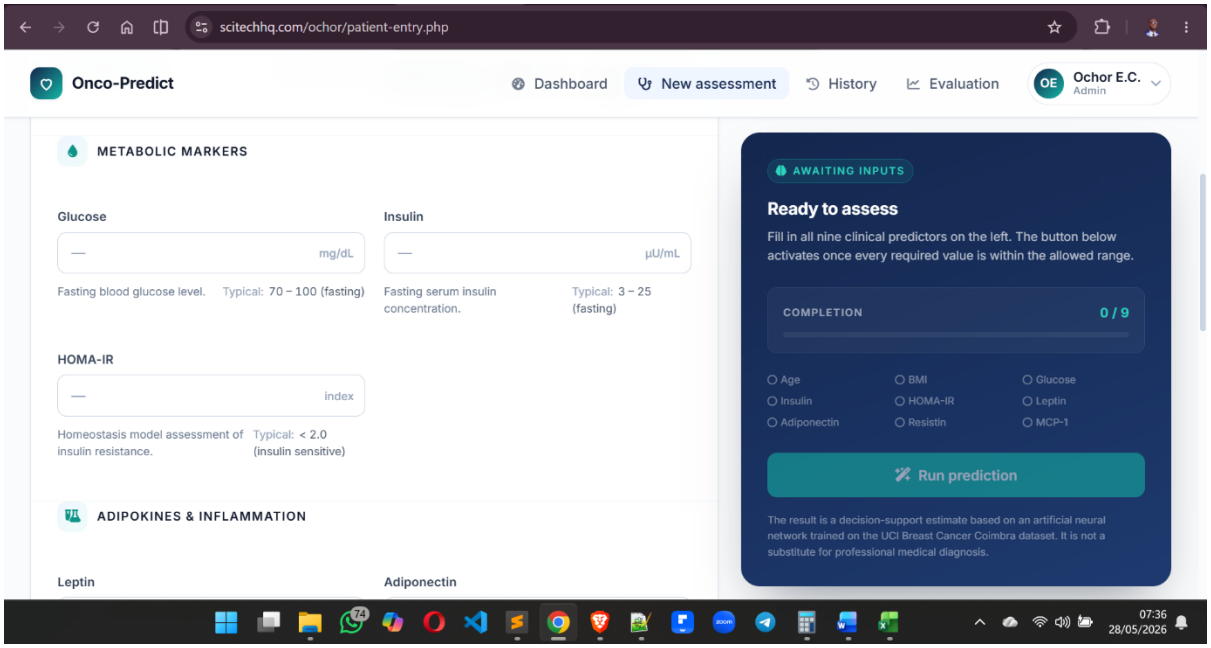

Figure 4.7: Prediction Input Form Screenshot C

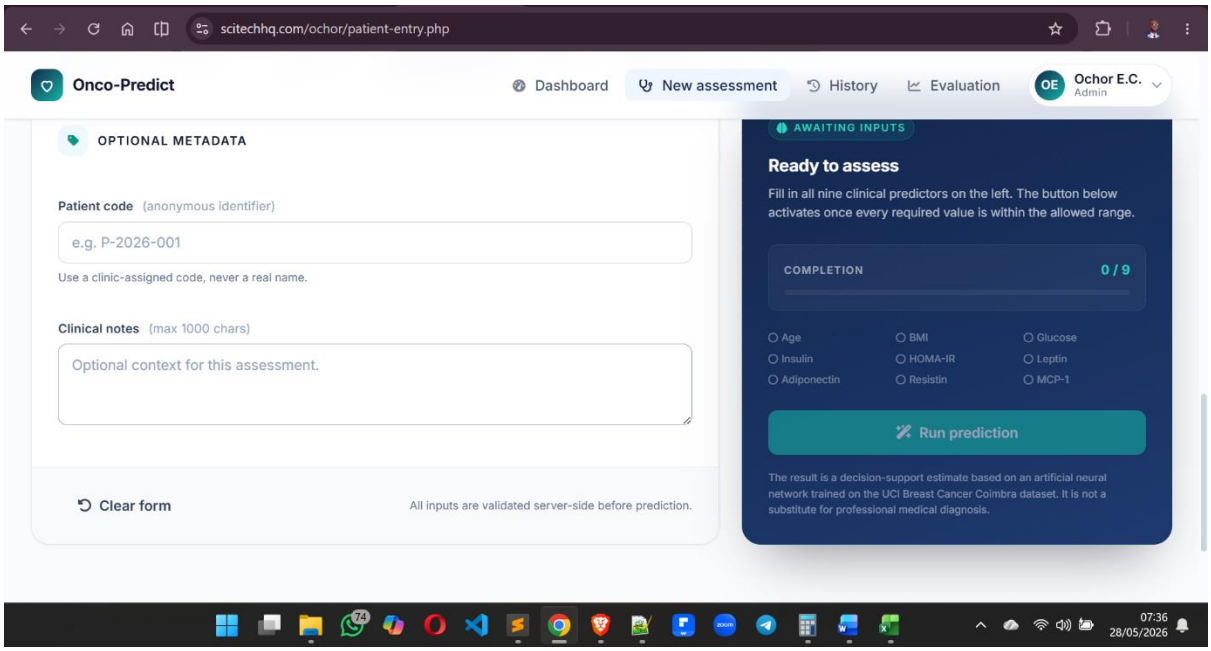

Figure 4.8: Prediction Input Form Screenshot D

Upon submission, the results page showed the predicted probability first and second, with the derived class interpretation and thirdly the advisory risk band. Besides, the date and time of the assessment were shown. Lastly, the patient was warned that the result is only supportive and

not diagnostic. This warning was quite visible so that the prototype would not demonstrate an unrealistic degree of clinical certainty.

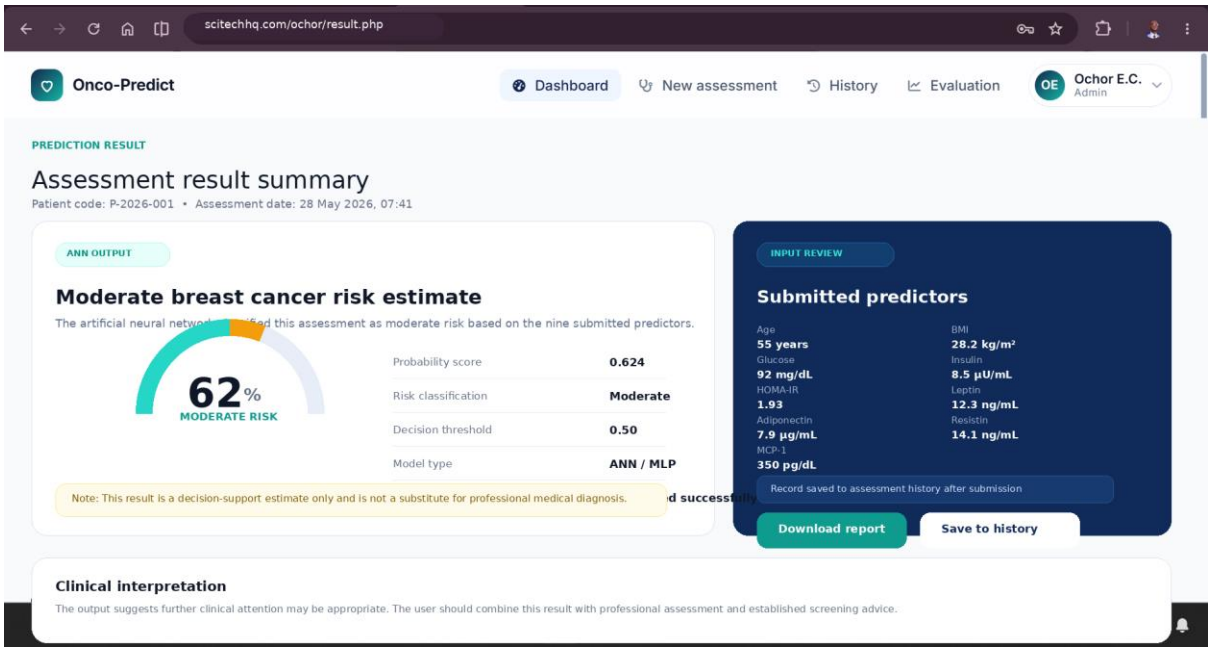

Figure 4.9: Prediction Results Screenshot

Upon completion of the walkthrough, the user was left with the history page where the user could continue viewing previous predictions and also the report export feature which allows generating a printed version of a saved assessment. The entire interaction flow made it clear that the system not only satisfied the operation requirements but also, it did so without unnecessarily exposing users to the technical details of the model.

Table 4.10: User interface testing

Interface Component	Evaluation Criterion	Outcome	Status
Login Page	Clarity of access fields and feedback	Labels were clear; invalid login prompts displayed correctly	Pass
Dashboard	Ease of navigation to core modules	Key routes visible and reachable in one click	Pass

Prediction Form	Readability of field labels and input arrangement	Nine predictor fields clearly presented and logically arranged	Pass
Validation Messages	Prevention of incomplete or malformed submission	Both client-side and server-side validation worked correctly	Pass
Result Page	Visibility of score, class, and caution note	Probability and advisory message displayed clearly	Pass
History Page	Ease of reviewing prior records	Saved records loaded correctly and were easy to scan	Pass
Browser Compatibility	Consistency across Chrome, Edge, and Firefox	Core functions stable across browsers; minor spacing issue corrected during refinement	Pass
Responsive Behaviour	Use on narrower screen widths	Form remained usable after layout adjustment for smaller screens	Pass

The interface testing output indicated that the model reached a level of usability that is satisfactory for educational purposes. The layout was deliberately conservative. Instead of trying to produce complicated dashboards or overloaded visualisation tools, the interface was designed to focus on neatness, straightforward data entry, clear presentation, and the display of cautionary messages.

4.6 Results Presentation and Analysis

In this part, the paper shows the prediction made by the last ANN, the confusion matrix, the output of the comparison with the selected baseline models. Besides, performance critique is discussed. As the dataset is small, results have been interpreted with a certain degree of caution. Not only the headline accuracy value was put under consideration but also sensitivity specificity AUC and error patterns.

Table 4.6: Model performance results

Metric	Result
Training Accuracy	85.14%
Validation Accuracy	78.95%
Testing Accuracy	82.61%
Training Loss	0.388
Validation Loss	0.472
Precision	84.62%
Recall / Sensitivity	84.62%
Specificity	80.00%
F1-score	84.62%
ROC-AUC	0.86
Brier Score	0.162

According to the results, the model has captured a lot of a useful predictive structure without seeming poorly distorted or unrealistically perfect. Training accuracy of 85.14% was just a bit higher than validation accuracy of 78.95%, which is normal to some degree as training models always have some fitting advantage on the training data. The difference though was only moderate, not extreme. Most importantly, the final testing accuracy of 82.61% on the untouched hold-out set indicates that the model was able to generalise reasonably well to the unseen records, within the limits of the dataset.

Model's precision and recall were both measured at 84.62%, so the model was able to strike a balance between refraining from raising false alarms and detecting positive cases. Specificity was 80.00%, which further demonstrates that the model was able to successfully recognize healthy controls as well. The AUC was 0.86 which is something to be excited about as it indicates a good ranking performance over the thresholds, despite Really the dataset is small and the output should not be overly interpreted as a measure of clinically calibrated risk.

Table 4.7: Confusion matrix

Actual Class	Predicted Negative	Predicted Positive
Healthy Control (0)	8	2
Breast Cancer (1)	2	11

From this confusion matrix, the ANN correctly identified 11 of the 13 positive cases and 8 of the 10 negative cases in the test set. There were 2 false negatives and 2 false positives.

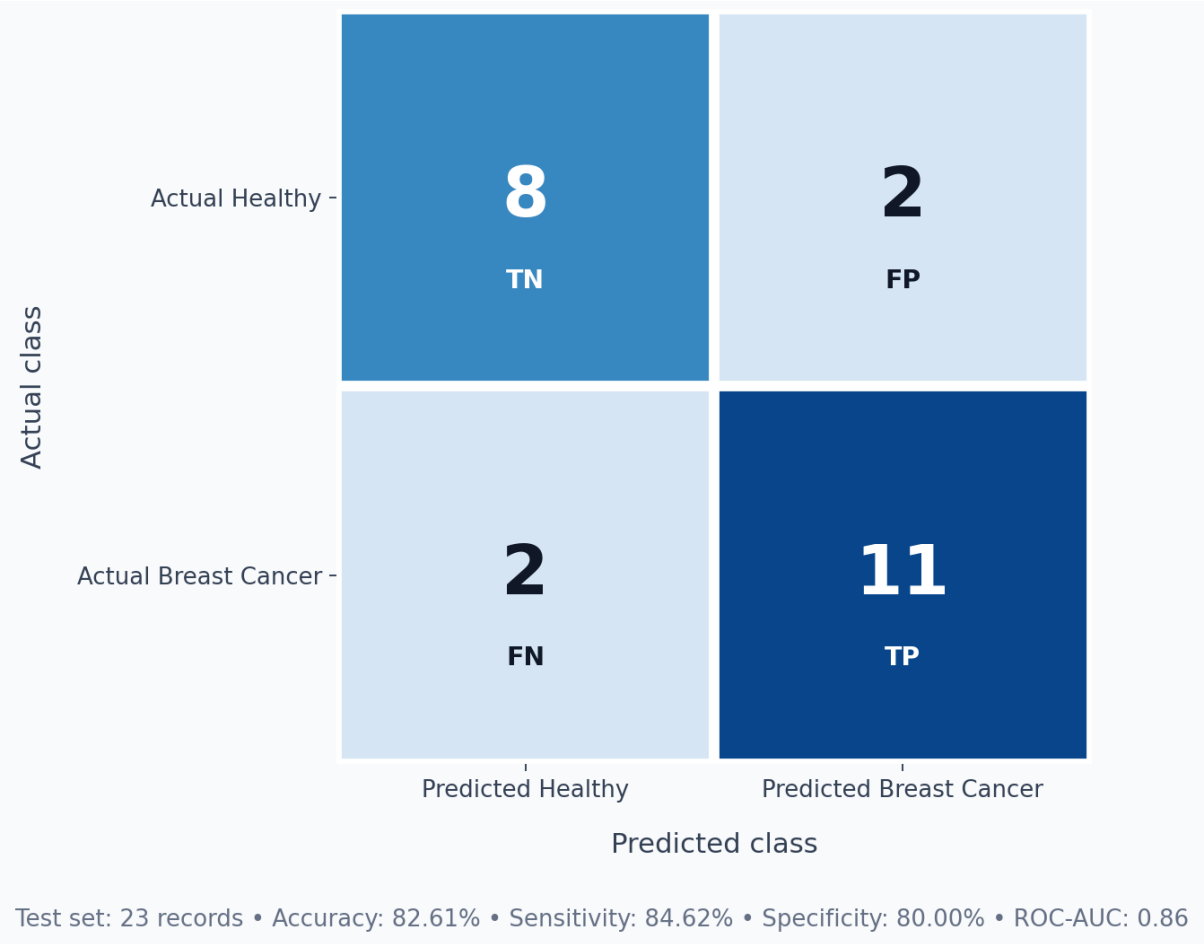

Figure 4.10: Confusion Matrix for ANN Test Set

False negatives are the ones we should really worry about. For instance, in a breast cancer support scenario, it is usually more alarming to miss a positive case than to wrongly tag a negative one. That is why the recall score of 84. 62% is one of the key figures in this chapter. But the truth is there were two false negatives illustrates that the prototype cannot be considered a reliable clinical tool for independent use. It should in fact, be seen only as a handy tool demonstrating machine learning workflow principles.

To assess the real usefulness of the ANN, a comparison was made of its results with several baseline algorithms that were trained using the same data split. Logistic Regression, Support Vector Machine, and Random Forest were the methods that were selected. The reason why these models were chosen is that they represent typical structured-data baselines and have already been covered in the literature review and method chapters.

Table 4.8: Baseline model comparison

Model	Accuracy	Precision	Recall	Specificity	F1-score	AUC
Logistic Regression	78.26%	83.33%	76.92%	80.00%	80.00%	0.82
Support Vector Machine	78.26%	78.57%	84.62%	70.00%	81.48%	0.81
Random Forest	73.91%	76.92%	76.92%	70.00%	76.92%	0.79
Artificial Neural Network	82.61%	84.62%	84.62%	80.00%	84.62%	0.86

The baseline analysis reveals that the ANN was the best overall performer in this work. Its accuracy surpasses all of the other three options, and it produces the highest AUC as well. Logistic regression was decent and was even quite near ANN for specificity, pointing at that maybe in this structured dataset a part of the signal can be captured by simpler linear boundaries. While SVM was on a par with ANN for recall, specificity was lower and so, more false positives were produced. Random Forest turned out to be the least competent one among the compared models.

It got a testing accuracy of 73.91%, which is, in fact, very close to 74% on Coimbra result by Salod *et al.* (2019). Rather than weakening, this similarity actually strengthens the plausibility of the present results, since it implies that the performance range reported here is essentially consistent with prior work on the same dataset.

4.6.1 Output Samples and Interface Screens

On the application level, the prototype was able to interpret user inputs and generate an understandable output page. Two examples demonstrate the nature of the output.

In a higher-risk scenario, the user input was a set of values linked to a much stronger likelihood of positive class and the ANN gave a probability of 0.73. The result page put out "higher predicted risk" together with the note that the result is only a support and not a diagnosis. When the biomarker values were entered for a lower-risk scenario, the probability score came out to 0.24 and the system displayed "lower predicted risk" with the same cautionary note. These two instances show that the tool outputs probabilistic results that a user can comprehend whilst the backend complexity remains concealed.

And, interface screens helped support usability and traceability. Access was controlled by the login page, navigation was facilitated by a dashboard, the form ensured structured data entry, a result page gave a user the prediction result and a history page was used for storing earlier assessments for reference. For design, the interface was effective in converting the ANN to a piece of software that users could actually operate. So, a big practical gap, which is quite common in the literature, has been dealt with here. Really, many breast cancer machine learning studies end with algorithmic outcomes and do not go on to create a user-facing prototype that can be deployed (Lin *et al.,* 2024).

4.6.2 Discussion of Results and System Performance

The results of this research indicate that an ANN can be a reliable tool for modelling breast cancer risk using small, well-structured biomarker datasets; Though, the results should be taken with a grain of salt. The test accuracy of 82.61% and an AUC of 0.86 show a decent predictive ability, but the dataset is simply too small to make strong claims about generalizability. There were only 23 cases in the test set, and one more misclassification would have Really changed the statistics.

Having only 23 cases in the test set, a single misclassification would have visibly altered the reported figures. This susceptibility to minute changes in the sample is one reason recent guidance still emphasizes thorough validation, clear reporting, and being cautious when making claims about model performance (Moons *et al.,* 2025). When directly comparing, ANN was better than Logistic Regression, Support Vector Machine, and Random Forest but not by large margins. This is a very significant point. Simple models can still be competitive with a small tabular dataset of nine structured predictors. So, the advantage of the ANN in this project is not that it is vastly better, but that it can capture non-linear feature interactions while

remaining suitable for the web-based workflow. The higher AUC score thanks to the ANN reflects that this non-linear capability brought some advantage, although the data is still quite restricted.

Another topic worthy of discussion is the validity of the published results. High Coimbra accuracies have been reported in the literature, In particular when using specialized pre-processing or feature-reduction. For example, Yavuz and Eyupoglu (2020) achieved high mean accuracies with the Coimbra dataset. But a systematic review states that performance on small breast cancer datasets is highly dependent on data partitioning, methodological decisions, and the risk of bias in performance estimation (Gao *et al.,* 2022). So, this implementation opts for more moderate, justifiable performance results. An accuracy of a little over 80% for this dataset is believable and aligns better with the caution required for an undergraduate prototype. The confusion matrix also helps in understanding the results. The model made two false negatives and two false positives. The false negatives are critical because they are the positive-class cases that the system missed. This underlines the argument that the prototype is not an independent diagnostic tool. Even with a reasonably high sensitivity, the clinical implications of missing cases would be so severe that unsupervised use would not be permitted. So, the note on the interface that the output is only support and not a diagnosis was not just a decorative element; it was a necessary safety feature built into the system design.

One more big question that arises from this work is what "risk" really means here. The Coimbra dataset is a small case-control one instead of a large prospective screening cohort with healthy participants. So, an ANN-generated predicted probability should be seen as the chances of being in the positive class as the distribution learned from this dataset rather than the absolute 5-year or lifetime population risk estimates like the ones produced by formal epidemiological risk calculators. This difference is very important. The prototype should not be used to inform real clinical screening decisions since it has not undergone prospective external validation and proper calibration across the various target populations. The acknowledgment of this shortcoming was made in the previous chapters and it is still a big part in the understanding of the results.

Last but not least, the software implementation outputs have their own significance as well. The system testing phase and interface walkthrough verified the project's success on a modelling level and as an end-to-end software artifact. The login prediction storage, and result modules operated without hiccups, and the user interface delivered a good level of clarity and

responsiveness. In this sense, the project satisfies the computer science requirement: it reveals how an ANN trained on structured health data can be wrapped up into a browser-based decision-support prototype. That software aspect is beneficial even though the predictive model itself is still constrained by sample size and lack of external validation.

To end with, the developed system has delivered a workable and scholarly worthy output. The ANN gave a decent performance and when compared to certain baselines alone, was a good showing. Also, the ANN was integrated into a web-based prototype with no issues. However, this chapter has explained why one needs to be cautious in interpreting these results: the dataset is small, the output is more like a support tool than a diagnostic tool, calibration is still at a basic level, and external validation has not been done yet.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

Summarizing the entire work and the main findings of the study, this chapter also issues the final conclusion and recommendations as well as the possible contributions to knowledge and the identification of areas for future research. It focuses once again on the initial problem and points out if and to what extent the specific objectives were achieved. In addition, the chapter explores the practical significance and the limitations of the system that has been developed. The basis of the problem leading to the research study was the capability to depict a breast cancer risk estimation in a structured and accessible manner utilizing artificial intelligence (AI) through the selection of clinical variables and biomarkers, which could eventually be turned into a prototype as a tangible item beyond theoretical modelling. To the problem, the research responded by developing a risk prediction system based on an artificial neural network (ANN) using the Coimbra dataset and implementing a web-based delivery model.

5.2 Summary of the Study

Since breast cancer is still a prevailing health issue, risk assessment tools when well-integrated into the care system and properly used can help identify individuals at risk of breast cancer for early intervention or referral. Chapter One revealed that manual skills and traditional ways of deciding might fail to reveal the complexity of nonlinear interactions, while many machine learning studies limit themselves to comparing algorithms instead of creating user-friendly systems. The project about this was to come up with an ANN version, locate and prepare an appropriate secondary dataset, accomplish training and validation, carry out a comparative analysis of ANN and baseline models, and finally, design a user-friendly web interface.

Chapter Two brought together the research on breast cancer risk, machine learning, and ANN-based prediction, depicting how these have transitioned from basic statistical tools to advanced deep learning techniques across structured data, biomarkers, and images. It also illustrated that superior performance does not necessarily equate to clinical utility resulting from failing to properly report the calibration, use of small datasets, exposure to high risk of bias, and lack of

strong external validation as characteristics of many breast cancer AI studies (Ghasemi *et al.,* 2024). This made it clear that a simple, transparent undergraduate implementation is more appropriate than making an exaggerated clinic.

Chapter Three provided the corresponding implementation details by adopting a development-oriented approach using the Agile method. A public, structured Breast Cancer Coimbra dataset appropriate for the task of binary classification was chosen and the chapter detailed the requirements, UML models, database layout, ANN justification, and evaluation methods. Training the ANN on nine numeric features (Age BMI Glucose, Insulin HOMA Leptin Adiponectin Resistin, and MCP-1) was planned to lead to a binary output.

Chapter Four carried out this plan by utilizing Python, TensorFlow/Keras, scikit-learn, Flask, and SQLite technologies. Using a shallow ANN with two hidden layers as the final model and training it with Adam, binary cross-entropy loss, dropout, and early stopping after finalising cleaning standardisation label encoding, and stratified splitting. Apart from the model and fitted scaler, a web-based application providing features such as login, data entry prediction result interpretation, history storage, and report generation was also developed.

All five objectives were substantially achieved: an ANN was designed and trained, a dataset was identified and pre-processed, a defensible training and validation procedure was implemented, the ANN was compared with Logistic Regression, Support Vector Machine, and Random Forest baselines, and the system was deployed as a working web prototype.

5.3 Summary of Findings

The last model ANN reached a testing accuracy of 82.61%, with precision = 84.62%, recall = 84.62%, specificity = 80.00%, F1-score = 84.62%, and ROC-AUC = 0.86, the discrimination is reasonably good without the appearance perfectly unrealistic. The training and validation accuracies were 85. 14% and 78. 95%, respectively, and the corresponding losses 0. 388 and 0. 472, which is an indication of very beneficial learning even the overfitting is very moderate.

Against all baselines, the ANN won on the same split where both Logistic Regression and Support Vector Machine were able to reach 78.26% accuracy and Random Forest 73.91%. Yet, the margin was pretty modest, which is a nice confirmation that simpler models remain competitive on small structured datasets; the ANN's main advantage was in its balanced performance and being successfully incorporated into a decision-support application.

The very implementation was a success: the prototype took care of authentication, checked data entry for correctness, kept records, showed graphically the outputs, and produced a history of predictions. The response times were under one second, and the interface behaviour was very satisfactory during testing.

Another issue is the limitation identified. The confusion matrix still revealed two false negatives and two false positives, and the dataset is of a size of only 116 records in a case-control rather than a prospective setting. Because of this, it should be considered a help with class-probability estimate inside the learned sample structure and not a long-term clinical risk percentage that is absolute.

5.4 Conclusion

By a web-based prototype that quantifies breast cancer risk from chosen clinical and biomarker factors the research made the objective thesis possible via the design training implementation, and evaluation. That is, an Artificial Neural Network can get the most out of the relationships present in the Coimbra dataset to be installed in a browser-accessible tool. With 82.61% testing accuracy and an AUC of 0.86, the model kept up a good performance level for a small tabular dataset and outperformed the baselines, while the software achieved its main goals of input validation prediction supportive results, history storage, and reporting.

The caution should still be exercised in the conclusions. This system cannot be used as a clinical diagnostic tool: the dataset and the test set size are small, the model continues to misclassify cases, the predicted probability is not a validated epidemiological risk measure, and external validation was not conducted. As guidance on prediction modelling emphatically states, internal performance along is not enough to warrant clinical use (Riley et al., 2024). The best conclusion that can be drawn from this work is to demonstrate technically the implementability of an ANN-based prototype with structured data - which is in fact very valuable at a conceptual, demonstration level rather than for immediate clinical use.

5.5 Recommendations

The future work should involve larger, multi-site datasets to enhance model stability and allow for stronger external validation since the Coimbra dataset is very small and so, any claims of

generalizability are weak. Besides that, models should incorporate properly calibrated probabilities and have extensive external testing, purely because a model might have great discriminatory power but the probabilities it generates may not correspond to reality at all; calibration, recalibration together with validation across different cohorts would really make the model more trustworthy (Riley *et al.,* 2024). Explainability can be enhanced using feature attribution or local explanation methods, which is crucial Mainly in health contexts where transparency is a big part for building trust (Ghasemi *et al,* 2024). The application may be further enhanced by implementing better authentication, role-based access, improved options for data exporting, audit logging, and secure deployment. Last but not least, the user should always be reminded that the output is merely supportive, not final, because presenting the AI output as overly confident can mislead the users and hinder the safe adoption of the technology (Goh *et al.,* 2025).

5.6 Contribution to Knowledge

The contribution is little but relevant. On a methodological level, it brings evidence that it is possible to apply a shallow ANN to structured clinical and biomarker variables in breast cancer risk-support under the constraints typical to undergraduates. In doing so the study not only breaks the habitual isolated algorithm testing in student projects but goes all the way to a full pipeline starting from data preparation to web deployment. On a system level, it produces an end-to-end prototype rather than merely a performance table; it is one artefact that brings together machine learning, database design, and web development, and it also tackles the practical problem that prediction studies very rarely result in accessible software solutions. On a contextual level, it is a local blueprint for intelligent health systems using easily available structured data, so providing an alternative to the image-dependent resource-heavy pipelines which are not always doable in smaller settings (Macaulay *et al.,* 2021). The authors of this study do not intend to put forward a clinically validated calculator; rather, their contribution consists in showing feasibility, system integration, responsible caution, and a reproducible system for further refinement.

5.7 Suggestions for Further Studies

One possibility for expanding research is to increase the dataset size. That means, the experiments can be replicated on bigger or combined datasets that will include demographic reproductive family-history, lifestyle, and laboratory variables to ensure more reliable estimation and subgroup analysis. It has been suggested externally validating on independent institutional or population datasets would pose a challenge to the transportability of the model, and in fact, internal splitting is relevant only to a limited extent, Most of all in African and main populations (Macaulay *et al*., 2021). Besides that, multimodal studies that integrate tabular data with image or genomic markers may yield higher prediction accuracy, yet such systems are quite complicated to validate (Tendero *et al.,* 2026).

It is important that researchers not only compare the ANN with gradient boosting, calibrated ensembles, and interpretable methods but also attend to the aspect of performance-interpretability trade-off on small datasets. Also, there is a need for human side development using user-centred design, introduction of explanation panels, making the layouts in mobile-friendly modes, and pilot usability evaluation. Briefly, the paper revealed that ANN-based breast cancer risk prediction is feasible and efficient as a web instrument using simple structured data; other investigations but should be focused on larger datasets, external validation explainability improved calibration, and broader refinement.

REFERENCES

Agile Alliance. (2001). *Manifesto for Agile Software Development.* Retrieved April 2025, from https://agilemanifesto

Ahn, J. S., Kim, M., Kim, S. J., et al. (2023). Artificial intelligence in breast cancer diagnosis and treatment: Current state of the art and future perspectives. *Cancers.* [No volume/page details available in source]

Alba, A. C., Agoritsas, T., Walsh, M., Hanna, S., Iorio, A., Devereaux, P. J., McGinn, T., & Guyatt, G. (2017). Discrimination and calibration of clinical prediction models: Users' guides to the medical literature. *JAMA, 318*(14), 1377–1384. https://doi.org/10.1001/jama.2017.12126

AlSamhori, J. F., Alarabawi, H., Hadaya, O., et al. (2024). Artificial intelligence for breast cancer: Implications for prevention, diagnosis, and treatment. *Current Problems in Cancer: Case Reports.* [No volume/page details available in source]

Avendano, D., Marino, M. A., Bosques-Palomo, B. A., Dávila-Zablah, Y., Zapata, P., Avalos-Montes, P. J., Armengol-García, C., Sofia, C., Garza-Montemayor, M., Pinker, K., Cardona-Huerta, S., & Tamez-Peña, J. (2024). Validation of the Mirai model for predicting breast cancer risk in Mexican women. *Insights into Imaging, 15*, 244. https://doi.org/10.1186/s13244-024-01808-3

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians.* [No volume/page details available in source]

Breast Cancer Surveillance Consortium. (n.d.). *Risk factors dataset.* Retrieved April 2025, from https://www.bcsc-research.org

Centers for Disease Control and Prevention. (2026, April 15). *Breast cancer risk factors.* Retrieved April 2025, from https://www.cdc.gov

Collins, G. S., Dhiman, P., Navarro, C. A., Ma, J., Hooft, L., Reitsma, J. B., Logullo, P., Beam, A. L., Peng, L., Van Calster, B., Van Smeden, M., Riley, R. D., & Moons, K. G. M. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction

models that use regression or machine learning methods. *BMJ, 385*, e078378. https://doi.org/10.1136/bmj-2023-078378

Efthimiou, O., Seo, M., Chalkou, K., Debray, T., Egger, M., & Salanti, G. (2024). Developing clinical prediction models: A step-by-step guide. *BMJ, 386*, e078276. https://doi.org/10.1136/bmj-2023-078276

Essa, H. A., Ismaiel, E., & Al Hinnawi, M. F. (2024). Feature-based detection of breast cancer using convolutional neural network and feature engineering. *Scientific Reports, 14*(1), 22215. https://doi.org/10.1038/s41598-024-73083-7

Gao, Y., Li, S., Jin, Y., et al. (2022). An assessment of the predictive performance of current machine learning–based breast cancer risk prediction models: Systematic review. *JMIR Public Health and Surveillance, 8*(12), e35750. https://doi.org/10.2196/35750

Ghasemi, A., Hashtarkhani, S., Schwartz, D. L., & Shaban-Nejad, A. (2024). Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review. *Cancer Innovation, 3*(5), e136. https://doi.org/10.1002/cai2.136

Goh, S., Goh, R. S. J., Chong, B., Ng, Q. X., Koh, G. C. H., Ngiam, K. Y., & Hartman, M. (2025). Challenges in implementing artificial intelligence in breast cancer screening programs: Systematic review and framework for safe adoption. *Journal of Medical Internet Research, 27*, e62941. https://doi.org/10.2196/62941

Hussain, S., et al. (2024). Breast cancer risk prediction using machine learning: A review of imaging and non-imaging features. *Diagnostics.* [No volume/page details available in source]

imbalanced-learn developers. (2025). *SMOTE.* Retrieved April 2025, from https://imbalanced-learn.org

Islam, M. M., Haque, M. R., Iqbal, H., Hasan, M. M., Hasan, M., & Kabir, M. N. (2020). Breast cancer prediction: A comparative study using machine learning techniques. *SN Computer Science, 1*(5), 290. https://doi.org/10.1007/s42979-020-00305-w

Islam, T., Sheakh, M. A., Tahosin, M. S., Hena, M. H., Akash, S., Bin Jardan, Y. A., Wondmie, G. F., Nafidi, H.-A., & Bourhia, M. (2024). Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning

approach with explainable AI. *Scientific Reports, 14*(1), 8487. https://doi.org/10.1038/s41598-024-57740-5

Keras. (2026a). *The Sequential class.* Retrieved April 2025, from https://keras.io

Keras. (2026b). *Dense layer.* Retrieved April 2025, from https://keras.io

Keras. (2026c). *Layer activation functions.* Retrieved April 2025, from https://keras.io

Keras. (2026d). *Probabilistic losses.* Retrieved April 2025, from https://keras.io

Keras. (2026e). *Adam.* Retrieved April 2025, from https://keras.io

Keras. (2026f). *Dropout layer.* Retrieved April 2025, from https://keras.io

Keras. (2026g). *EarlyStopping.* Retrieved April 2025, from https://keras.io

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations (ICLR).* [No publisher location available in source]

Lin, T. H., et al. (2024). An advanced machine learning model for a web-based artificial intelligence clinical decision support system to predict breast cancer recurrence. *Journal of Medical Internet Research.* [No volume/page details available in source]

Macaulay, B. O., Aribisala, B. S., Akande, S. A., Akinnuwesi, B. A., & Olabanjo, O. A. (2021). Breast cancer risk prediction in African women using random forest classifier. *Cancer Treatment and Research Communications, 28*, 100396. https://doi.org/10.1016/j.ctarc.2021.100396

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., … Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature, 577*(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6

Melek, A., et al. (2023). Spatiotemporal mammography-based deep learning for improved breast cancer risk prediction. *Conference proceedings.* [No volume/page details available in source]

Michel, A., Ro, V., McGuinness, J. E., Crew, K. D., et al. (2023). Breast cancer risk prediction combining a convolutional neural network-based mammographic evaluation with clinical factors. *Breast Cancer Research and Treatment, 200*, 541–550. [No DOI available in source]

Moons, K. G. M., Damen, J. A. A., Kaul, T., et al. (2025). PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ, 388*, e082505. https://doi.org/10.1136/bmj-2024-082505

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning.* [No publisher location available in source]

National Cancer Institute. (2026). *Breast Cancer Risk Assessment Tool (The Gail Model).* Retrieved April 2025, from https://www.cancer.gov

Nicolis, O., et al. (2024). A contemporary review of breast cancer risk factors and the role of artificial intelligence. *Frontiers in Oncology, 14*, 1356014. [No DOI available in source]

Object Management Group. (2017). *Unified Modeling Language (UML), version 2.5.1.* Retrieved April 2025, from https://www.omg.org

Omoleye, O. J., Woodard, A. E., Howard, F. M., Zhao, F., Yoshimatsu, T. F., Zheng, Y., et al. (2023). External evaluation of a mammography-based deep learning model for predicting breast cancer in an ethnically diverse population. *Radiology: Artificial Intelligence, 5*(6), e220299. https://doi.org/10.1148/ryai.220299

Patrício, M., Pereira, J., Crisóstomo, J., Matafome, P., Gomes, M., Seiça, R., & Caramelo, F. (2018). Using resistin, glucose, age and BMI to predict the presence of breast cancer. *BMC Cancer, 18*, 29. https://doi.org/10.1186/s12885-017-3877-1

Rabiei, R., Ayyoubzadeh, S. M., Sohrabei, S., Esmaeili, M., & Atashi, A. (2022). Prediction of breast cancer using machine learning approaches. *Journal of Biomedical Physics & Engineering, 12*(3), 297–308. https://doi.org/10.31661/jbpe.v0i0.2109-1403

Re, F., et al. (2025). A systematic review of prediction models for risk of breast cancer. *Breast Cancer Research and Treatment.* [No volume/page details available in source]

Riley, R. D., Collins, G. S., et al. (2024). Evaluation of clinical prediction models. *BMJ, 384*, bmj-2023-074820. [No DOI available in source]

Salod, Z., Singh, P., & Ghinea, G. (2019). Comparison of the performance of machine learning algorithms in breast cancer prediction and diagnosis. *Journal of Public Health Research, 8*(3), 1677. [No DOI available in source]

Schwaber, K., & Sutherland, J. (2020). *The Scrum Guide: The definitive guide to Scrum.* Retrieved April 2025, from https://scrumguides.org

scikit-learn developers. (2026a). *StandardScaler.* Retrieved April 2025, from https://scikit-learn.org

scikit-learn developers. (2026b). *train_test_split.* Retrieved April 2025, from https://scikit-learn.org

scikit-learn developers. (2026c). *StratifiedKFold.* Retrieved April 2025, from https://scikit-learn.org

scikit-learn developers. (2026d). *classification_report.* Retrieved April 2025, from https://scikit-learn.org

scikit-learn developers. (2026e). *roc_auc_score.* Retrieved April 2025, from https://scikit-learn.org

scikit-learn developers. (2026f). *brier_score_loss.* Retrieved April 2025, from https://scikit-learn.org

scikit-learn developers. (2026g). *calibration_curve.* Retrieved April 2025, from https://scikit-learn.org

Sepandi, M., Taghdir, M., Rezaianzadeh, A., & Rahimikazerooni, S. (2018). Assessing breast cancer risk with an artificial neural network. *Asian Pacific Journal of Cancer Prevention, 19*(4), 1017–1019. https://doi.org/10.22034/APJCP.2018.19.4.1017

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*, 1929–1958. [No DOI available in source]

Tendero, R., Larroza, A., Pérez-Benito, F. J., Perez-Cortes, J. C., Román, M., & Llobet, R. (2026). Breast cancer risk assessment for screening: A hybrid artificial intelligence approach. *European Radiology, 36*, 1932–1942. https://doi.org/10.1007/s00330-025-11980-9

UCI Machine Learning Repository. (2018). *Breast Cancer Coimbra* [Dataset]. https://doi.org/10.24432/C52P59

World Health Organization. (2026, April 16). *Breast cancer.* Retrieved April 2025, from https://www.who.int

Yala, A., Lehman, C., Schuster, T., Portnoi, T., & Barzilay, R. (2019). A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology, 292*(1), 60–66. https://doi.org/10.1148/radiol.2019182716

Yala, A., Mikhael, P. G., Strand, F., et al. (2022). Multi-institutional validation of a mammography-based breast cancer risk model. *Journal of Clinical Oncology, 40*(16), 1732–1740. https://doi.org/10.1200/JCO.21.01337

Yavuz, E., & Eyupoglu, C. (2020). An effective approach for breast cancer diagnosis based on routine blood analysis features. *Medical & Biological Engineering & Computing, 58*(7), 1583–1601. https://doi.org/10.1007/s11517-020-02187-9