ÇUKUROVA ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

DOKTORA TEZİ

Breast Cancer Classification using Effective Machine Learning Techniques

Nagham Rasheed Hameed ALSAEDI

Bilgisayar Mühendisliği Anabilim Dalı

Kasım, 2025

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DOKTORA TEZ ONAYI

Breast Cancer Classification using Effective Machine Learning Techniques

---

Nagham Rasheed Hameed ALSAEDI

Bilgisayar Mühendisliği Anabilim Dalı

Bu Doktora Tezi 13/11/2025 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından Değerlendirilmiş ve ~~Oy Birliği~~ / Oy Çokluğu ile Kabul Edilmiştir.

Jüri	: Prof. Dr. M. Fatih AKAY (Advisor)	……
	: Prof. Dr. M. Uğraş CUMA	……
	: Prof. Dr. İbrahim Taner OKUMUŞ	……
	: Assoc. Prof. Dr. Serkan KARTAL	……
	: Assoc. Prof. Dr. Fatih KILIÇ	……

Bu Tez Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalında Hazırlanmıştır.
Tez No:

Prof. Dr. Sadık DİNÇER
Enstitü Müdürü

İÇİNDEKİLER

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DOKTORA TEZİ

Breast Cancer Classification using Effective Machine Learning Techniques

Nagham Rasheed Hameed ALSAEDI

Danışman: Prof. Dr. M. Fatih AKAY

Bilgisayar Mühendisliği Anabilim Dalı

ÖZ

Meme kanseri, dünya genelinde kadınlar arasında önde gelen ölüm nedenlerinden biri olmaya devam etmektedir ve bu durum pratik tanı araçlarına olan acil ihtiyacı ortaya koymaktadır. Bu çalışmada, meme kanseri sınıflandırma doğruluğunu artırmak amacıyla geliştirilmiş bir makine öğrenimi algoritması sunulmaktadır. Sistem, özellik çıkarımı için derin çok katmanlı algılayıcı (Deep MLP), boyut indirgeme için özellik birleştirilmiş bir otomatik kodlayıcı (feature-fused autoencoder) ve kare ağırlık ayarlaması ile çapraz doğrulama kullanılarak optimize edilmiş ağırlık ayarlı bir karar ağacı sınıflandırıcısını entegre etmektedir. Önerilen yöntemin sonuçları, sağlamlık ve genellenebilirliği sağlamak amacıyla Wisconsin meme kanseri veri kümesi üzerinde k-katlı çapraz doğrulama kullanılarak test edilmiştir. Ayrıca, gizli katman birimi sayısı, bırakma oranı (dropout rate), yığın boyutu (batch size) ve eğitim-test yüzdeleri gibi çeşitli hiperparametreler altında optimizasyon yapılmıştır. Tüm bu koşullar altında modelin performansı; doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1-skoru ve eğri altı alan (AUC) gibi temel metriklerle değerlendirilmiştir. Elde edilen sonuçlar, bu yaklaşımın geleneksel sınıflandırma yöntemlerinden daha iyi performans gösterdiğini ve veri bölümleri arasında yüksek doğruluk ve güçlü bir performans sergilediğini ortaya koymaktadır. Bu araştırmanın temel katkısı, derin öğrenme, otomatik kodlayıcı ve karar ağacı yöntemlerini birleştiren yeni bir çerçevenin geliştirilmesi olup, sistemin meme kanseri teşhisinde güçlü bir potansiyele sahip olduğunu ve hekimlere güvenilir, etkili bir araç sunduğunu göstermektedir. Ayrıca, geleneksel meme kanseri veri setlerine göre çok daha büyük ve çeşitli olan METABRIC veri setinin kullanılması, önerilen hibrit modelin değerlendirilmesi için güçlü bir test ortamı sağlamıştır. Klinik ve genomik özelliklerin zenginliği, modelin genelleme kapasitesinin daha kapsamlı bir şekilde doğrulanmasına olanak tanımış ve karmaşık gerçek dünya verilerini işleme konusundaki etkinliğini ortaya koymuştur.

Anahtar Kelimeler: Meme kanseri, Derin MLP, Özellik birleştirilmiş otomatik kodlayıcı, Ağırlık ayarlı karar ağacı.

ÇUKUROVA UNIVERSITY

INSTITUTE OF NATURAL AND APPLIED SCIENCES

PhD THESIS

Breast Cancer Classification using Effective Machine Learning Techniques

---

Nagham Rasheed Hameed ALSAEDI

Supervisor: Prof. Dr. M. Fatih AKAY

Department of Computer Engineering,

ABSTRACT

Breast cancer remains a leading cause of death among women worldwide, underscoring the urgent need for practical diagnostic tools. This work presents an advanced machine learning algorithm designed to enhance the classification accuracy of breast cancer. The system integrates a deep multi-layer perceptron (Deep MLP) for feature extraction, a feature-fused autoencoder for efficient dimensional reduction, and a weight-tuned decision-tree classifier optimized by cross-validation and square weight adjustment. The Wisconsin breast cancer dataset is utilized to test the results of the method rigorously using k-fold cross-validation. Optimizing performance has been done under different hyperparameters, namely, the number of hidden units, dropout rate, batch size, as well as test-train percentages. The performance of the model was evaluated under all the given conditions for key Metrics. These are the Accuracy, Precision, Recall, F1-score, and area under the curve (AUC). Using the above metrics, the model was able to distinguish malignant and benign tumors. Our findings show that this approach performs better than traditional classification methods, leading to accurate and robust results across several data partitions. This research contributes to a new framework pertaining to deep learning, Auto-encoder, and decision tree, which clearly shows that this framework has a very high probability of impacting breast cancer diagnosis while providing usefulness to physicians.

Furthermore, the use of the METABRIC dataset ,which is substantially larger and more diverse than classical breast cancer datasets ,provided a robust evaluation environment for the proposed hybrid model. Its rich clinical and genomic features enabled deeper validation of the model's generalization capability and demonstrated its effectiveness in handling complex, real-world data.

Keywords: Breast cancer, Deep MLP, feature fused autoencoder, weight-tuned decision tree.

GENİŞLETİLMİŞ ÖZET

Meme kanseri, dünya çapında kadınlar arasında önde gelen morbidite ve mortalite nedenlerinden biridir ve tanısal görüntüleme, moleküler profilleme ve hedefli tedavilerdeki önemli ilerlemelere rağmen önemli bir küresel sağlık sorunu oluşturmaya devam etmektedir. Meme tümörlerinin erken teşhisi ve doğru sınıflandırılması, tedavi başarısını artırmada, tekrarlama oranlarını azaltmada ve genel hasta sağkalımını iyileştirmede önemli bir rol oynamaktadır. Mamografi, ultrason ve histopatolojik değerlendirme gibi geleneksel tanı yaklaşımları, malign lezyonların belirlenmesinde altın standart olmaya devam etmektedir. Ancak bu yöntemler genellikle emek yoğun, zaman alıcıdır ve radyologların ve patologların öznel yorumlarına büyük ölçüde bağımlıdır. İnsan yargısındaki değişkenlik ve klinik uzmanlıktaki farklılıklar, tanısal tutarsızlıklara yol açarak hasta yönetimini geciktirebilir veya yanlış yönlendirebilir. Bu bağlamda, yapay zekâ (YZ), özellikle makine öğrenimi (ML) ve derin öğrenme (DL), tıbbi görüntüleme ve tanı otomasyonunda dönüştürücü bir araç olarak ortaya çıkmıştır. Bu hesaplamalı yaklaşımlar, sistemlerin verilerden otomatik olarak öğrenmesini, insan gözüyle görülemeyen ince, yüksek boyutlu örüntüleri belirlemesini ve tutarlı, tekrarlanabilir ve son derece doğru tanı çıktıları üretmesini sağlar. Sonuç olarak, Makine Öğrenmesi ve Öğrenmenin (ML) klinik iş akışlarına entegrasyonu, yalnızca verimliliği artırmakla kalmaz, aynı zamanda hekimlere erken ve kesin hastalık sınıflandırmasında yardımcı olabilecek gelişmiş klinik karar destek sistemlerinin temelini oluşturur.

Bu çalışmanın temel amacı, geleneksel istatistiksel veya kural tabanlı yaklaşımlara kıyasla üstün doğruluk, sağlamlık ve genelleştirilebilirlik gösteren bir meme kanseri sınıflandırma çerçevesi tasarlamak ve geliştirmektir. Önerilen çerçeve, her biri sınıflandırma sürecinde belirli bir aşamayı ele alan üç sinerjik bileşeni bir araya getirir.

Bunlar şunlardır:

1. Ham girdi verilerinden özellik çıkarımı için kullanılan Derin Çok Katmanlı Algılayıcı (Derin Çok Katmanlı Algılayıcı).
2. Boyut azaltma, yedeklilik giderme ve bilgi sıkıştırmadan sorumlu Özellik Kaynaştırılmış Otomatik Kodlayıcı .
3. Veri dengesizliğini telafi ederken nihai sınıflandırmayı gerçekleştiren Ağırlık Ayarlı Karar Ağacı. Bu üçlü tasarım, derin öğrenmenin temsil gücünü, ağaç tabanlı modellerin yorumlanabilirliği ve sınıf duyarlılığıyla birleştirmeyi amaçlamaktadır.

Derin MLP, özellikle girdi nitelikleri arasındaki yüksek düzeyli, doğrusal olmayan etkileşimleri yakalar; otokodlayıcı, bu öğrenilmiş temsilleri kompakt, bilgi yoğun özelliklere

entegre eder ve iyileştirir; ağırlık ayarlı karar ağacı ise çoğunluk sınıfına yönelik önyargıyı azaltmak için uyarlanabilir sınıf ağırlıklandırmasını kullanır. Bu modüller birlikte, hem sinir ağlarının tahmin gücünden hem de sembolik karar almanın şeffaflığından yararlanan tutarlı bir hibrit çerçeve oluşturarak, klinik kabul için gerekli özellikler olan doğru, kararlı ve yorumlanabilir bir sınıflandırıcı üretir.

Bu çalışmada kullanılan veri kümesi, literatürde meme kanseri tahmin modellerini değerlendirmek için yaygın olarak kullanılan, iyi bilinen bir kıyaslama veri kümesi olan Wisconsin Meme Kanseri Veri Kümesi'dir (WBCD). Meme kitlelerinin ince iğne aspirasyonu (İİA) görüntülerinden elde edilen hücre çekirdeklerinin ortalama yarıçapı, dokusu, çevresi, alanı, pürüzsüzlüğü, sıkılığı ve içbükeyliği gibi istatistiksel ölçümler de dahil olmak üzere 32 nicel özellikle karakterize edilen 569 hasta örneğinden oluşmaktadır. Veri ön işleme, model güvenilirliğini ve performans tutarlılığını sağlamak için önemli bir ön adımdı. Eksik veya anormal değerler, yükleme ve filtreleme prosedürleri yoluyla kaldırıldı.

Değişkenler arasındaki büyüklük farklılıklarını ortadan kaldırmak ve böylece yüksek değerli niteliklerin baskınlığını önlemek için tüm özellikler standart ölçekleme teknikleri kullanılarak normalleştirildi. Genelleştirilebilirliği değerlendirmek için, farklı veri kullanılabilirliği senaryoları altında sağlamlık değerlendirmesine olanak tanıyan %90-10, %80-20, %70-30 ve %55-45 gibi çoklu eğitim-test bölmeleri kullanıldı. Ayrıca, örnekleme yanlılığını azaltmak için k katlı çapraz doğrulama (tipik olarak k = 10) uygulandı ve her örneğin hem eğitim hem de doğrulama süreçlerine katkıda bulunduğundan emin olundu. Modelin katman başına gizli birim sayısı, bırakma olasılığı, toplu boyut, öğrenme oranı ve aktivasyon fonksiyonu gibi hiperparametreleri, en yüksek performans ve yakınsama kararlılığı için en uygun yapılandırmayı belirlemek amacıyla sistematik ızgara araması ve doğrulama izleme yoluyla ince ayarlandı.

Model mimarisi, karmaşıklık ve yorumlanabilirlik arasında denge kurmak için dikkatlice tasarlanmıştır. Deep MLP modülü, her biri çok seviyeli veri soyutlamalarını yakalamak için farklı sayıda nöronla donatılmış, birden fazla gizli katmana sahip yüksek kapasiteli bir özellik çıkarıcı olarak işlev görür. ReLU aktivasyon fonksiyonu, hesaplama verimliliği ve geri yayılım sırasında kaybolan gradyan sorunlarını azaltma yeteneği nedeniyle seçilmiştir ve böylece derin hiyerarşik öğrenmeye olanak sağlamıştır. Ağ, uyarlanabilir öğrenme hızı ve geleneksel gradyan iniş yöntemlerine göre üstün yakınsama özellikleri nedeniyle seçilen ADAM optimize edicisi kullanılarak eğitilmiştir. WBCD gibi orta büyüklükteki veri kümeleri üzerinde eğitilen derin ağlarda yaygın bir zorluk olan aşırı uyumu önlemek için, bırakma, toplu normalleştirme ve erken durdurma gibi düzenleme teknikleri kullanılmıştır. Daha sonra gelen Feature-Fused Autoencoder modülü ikili bir rol oynamıştır: çıkarılan özellikleri daha düşük boyutlu gizli bir gösterime sıkıştırmak ve önemli tanı bilgilerini korumak için bunları yeniden yapılandırmak. Standart otokodlayıcıların aksine, önerilen birleşik varyant, hem düşük hem de yüksek seviye özellik hiyerarşilerini birleştiren bir mekanizma içerir ve bu da daha ayırt edici kodlamayı mümkün kılar

ve yedekliliği azaltır. Bu, boyutsallığın azaltılmasının klinik açıdan ilgili örüntüleri tehlikeye atmamasını sağlar. Son olarak, Ağırlık Ayarlı Karar Ağacı sınıflandırıcısı, karar verme katmanı olarak işlev görür. Sınıf frekanslarına orantılı uyarlanabilir sınıf ağırlıkları atayarak, kötü huylu örneklerin genellikle yetersiz temsil edildiği tıbbi veri kümelerinde yaygın bir sorun olan sınıf dengesizliğini etkili bir şekilde telafi eder. K-kat doğrulaması yoluyla, en kararlı ve doğru ağaç yapısı seçilmiş ve yorumlanabilir ve dengeli sınıflandırma sınırları elde etmek için Gini safsızlığı ve bilgi kazanımı gibi bölünmüş ölçütler optimize edilmiştir.

Sistemin etkinliğini kapsamlı bir şekilde değerlendirmek için Doğruluk, Hassasiyet, Geri Çağırma (Hassasiyet), F1 puanı ve Alıcı İşletim Karakteristik Eğrisi Altındaki Alan (ROC-AUC) dahil olmak üzere çeşitli performans ölçütleri kullanılmıştır. Bu metrikler, öngörü kalitesinin çeşitli yönlerini birlikte değerlendirir; doğruluk genel doğruluğu niceliksel olarak belirler; kesinlik pozitif öngörülerin güvenilirliğini ölçer; hatırlama kötü huylu vakaları doğru bir şekilde tespit etme yeteneğini değerlendirir; ve F1 puanı kesinlik ve hatırlama arasında denge sağlayan harmonik bir ortalama sağlar. Özellikle ROC-AUC metriği, çeşitli sınıflandırma eşikleri boyunca ayırt etme yeteneğini değerlendirir. Deneysel sonuçlar, önerilen çerçevenin temel ve geleneksel modellerden önemli ölçüde daha iyi performans gösterdiğini ortaya koymuştur. Sistem %99,47 doğruluk, %99 kesinlik, %99 hatırlama, %99 F1 puanı ve 0,996'lık bir AUC elde etti. Gerçek Pozitifler (TP) = 350, Yanlış Negatifler (FN) = 7, Gerçek Negatifler (TN) = 208 ve Yanlış Pozitifler (FP) = 4'ten oluşan karışıklık matrisi, duyarlılık ve özgüllük arasında güçlü bir denge göstermektedir. Bu, modelin hem neredeyse tüm pozitif kanser vakalarını tespit edebildiğini hem de çok düşük bir yanlış pozitif oranını koruyabildiğini gösteriyor; yanlış sınıflandırmanın ciddi klinik sonuçlara yol açabileceği tıbbi teşhislerde önemli bir özellik.

Olağanüstü performans ve yüksek AUC değerleri, derin özellik öğrenme, optimize edilmiş boyut azaltma ve uyarlanabilir sınıf ağırlıklandırmasının sinerjik entegrasyonunun sistemin tanı gücünü önemli ölçüde artırdığını göstermektedir. Derin MLP bileşeni, biyomedikal özellikler arasındaki karmaşık, doğrusal olmayan bağımlılıkları etkili bir şekilde yakalayarak verilerin zengin bir gizli temsilini sağlar. Özellik kaynaştırılmış otokodlayıcı, gürültüyü ve alakasız bilgileri filtreleyerek bu temsilleri daha da iyileştirerek özlü ancak bilgilendirici bir özellik alanı oluşturur. Ağırlık ayarlı karar ağacı daha sonra güvenilir ve klinik olarak anlaşılabilir sınıflandırma sonuçları üretmek için yorumlayıcı mantığı uygular. Bu hibrit kombinasyon, yalnızca doğruluğu en üst düzeye çıkarmakla kalmaz, aynı zamanda karar şeffaflığını da korur; bu, yapay zeka destekli tanı araçlarının klinik olarak benimsenmesi ve güvenilmesi için temel bir gerekliliktir. Sonuç olarak, önerilen çerçeve, öngörücü performans ve yorumlanabilirlik arasındaki boşluğu kapatma kapasitesini göstererek, klinik olarak uygulanabilir akıllı tanı sistemlerine giden pratik bir yol sunmaktadır.

Umut verici sonuçlara rağmen, çalışma gelecekteki araştırmalarda ele alınması gereken bazı sınırlamaları da kabul etmektedir. WBCD veri kümesi, yaygın olarak doğrulanmış ve iyi

açıklamalı olmasına rağmen, nispeten küçük ve homojen bir örneklem kümesini temsil etmektedir. Gerçek dünyadaki klinik popülasyonlarda gözlemlenen karmaşıklığı ve heterojenliği tam olarak yansıtamayabilir. Bu nedenle, modelin klinik güvenilirliğe ulaşması için, mamografi, ultrason, MRI, histopatoloji ve gen ifadesi verileri gibi çeşitli görüntüleme ve genomik yöntemleri kapsayan daha büyük, çok merkezli ve çok modlu veri kümelerinde daha ileri düzeyde doğrulanması gerekmektedir. Bir diğer kritik husus ise modelin yorumlanabilirliğidir. Karar ağacı bileşeni bir miktar şeffaflık sağlarken, derin sinir bileşenleri (MLP ve otokodlayıcı) doğası gereği kara kutu modelleri olarak işlev görür. Klinik entegrasyon için hekimlerin kararların nasıl elde edildiğini izleyebilmeleri ve anlayabilmeleri gerekir. Bu nedenle, SHAP (SHapley Eklemeli Açıklamalar), LIME (Yerel Yorumlanabilir Modelden Bağımsız Açıklamalar) veya dikkat tabanlı görselleştirme mekanizmaları gibi açıklanabilir yapay zekâ (XAI) tekniklerinin dahil edilmesi, yorumlanabilirliği artırarak klinisyenler tarafından güven ve kabul görmeyi kolaylaştırabilir.

Bu çalışmanın pratik katkıları üç ana noktada özetlenebilir. İlk olarak, önerilen çerçeve, geleneksel makine öğrenmesi algoritmalarına kıyasla önemli bir performans artışı sağlamakta ve bu da tıbbi veri analizinde derin hiyerarşik temsil öğreniminin faydasını göstermektedir. İkinci olarak, derin sinirsel ve sembolik ağaç tabanlı modelleri birleştirerek, yüksek doğruluk ile yorumlanabilirlik arasında başarılı bir denge kurmaktadır; bu, klinik ortamlar için kritik bir gerekliliktir. Üçüncü olarak, bir sınıf ağırlıklandırma mekanizmasının dahil edilmesi, biyomedikal verilerdeki en yaygın sorunlardan biri olan sınıf dengesizliğini etkili bir şekilde ele alarak azınlık (kötü huylu) vakaların göz ardı edilmemesini sağlar. Bu katkılar bir araya geldiğinde, çerçeveyi tıp uzmanlarına tanısal akıl yürütme ve risk sınıflandırmasında yardımcı olabilecek güvenilir bir klinik karar destek sistemine dönüştürmek için sağlam bir temel oluşturur.

Geleceğe bakıldığında, mevcut çalışmayı genişletmek için çeşitli araştırma alanları öngörülmektedir. Gelecekteki çalışmalar, klinik çeşitliliği ve veri heterojenliğini daha iyi yansıtan büyük ölçekli, çok kurumlu veri kümeleri kullanarak modeli doğrulamaya odaklanacaktır. Histopatoloji slaytları, radyografik görüntüleme ve moleküler belirteçler gibi çok modlu veri kaynaklarının entegre edilmesi, çerçevenin meme kanserinin daha kapsamlı ve bağlam-farkında temsillerini öğrenmesini sağlayacaktır. Ayrıca, hafif mimariler, model budama ve niceleme teknikleri aracılığıyla hesaplama verimliliğinin optimize edilmesi, klinik iş akışlarında gerçek zamanlı uygulamayı destekleyecektir. Açıklanabilirliğin artırılması, tahminlerin şeffaflığını sağlamak ve hekimler tarafından incelenebilecek karar gerekçeleri sağlamak için XAI tekniklerinin daha fazla entegre edilmesiyle temel bir öncelik olmaya devam edecektir. Son olarak, transfer öğrenme, topluluk öğrenme ve meta-öğrenme gibi gelişmiş stratejilerin araştırılması, modelin uyarlanabilirliğini ve daha önce görülmemiş veri kümelerine genelleştirilmesini artırarak sistemi daha dayanıklı ve klinik olarak güvenilir hale getirebilir.

Sonuç olarak, bu çalışma, meme kanseri sınıflandırması için Derin Çok Katmanlı Öğrenme (MLP), özelliklerle birleştirilmiş bir otokodlayıcı ve ağırlık ayarlı bir karar ağacının güçlü yönlerini

birleştiren sağlam ve hibrit bir makine öğrenimi çerçevesi sunmaktadır. Önerilen yaklaşım, olağanüstü yüksek doğruluk ve güvenilirlik sağlayarak gerçek dünya teşhis uygulamaları için potansiyelini vurgulamaktadır. Karmaşık veri ilişkilerini etkili bir şekilde yakalayarak, yedekliliği azaltarak ve sınıf dengesizliğini ele alarak, model otomatik meme kanseri tespiti için kapsamlı bir çözüm sunmaktadır. Elde edilen sonuçlar, yapay zeka destekli erken teşhis ve karar desteğine doğru önemli bir adımı temsil etmektedir. Büyük ölçekli veri kümeleri üzerinde daha fazla doğrulama ve geliştirilmiş açıklanabilirlik mekanizmalarıyla, önerilen sistem araştırmadan gerçek dünya klinik uygulamasına geçiş için güçlü bir potansiyele sahiptir ve erken teşhise ve iyileştirilmiş hasta sonuçlarına anlamlı bir şekilde katkıda bulunmaktadır.

Bu çalışmada METABRIC veri setinin kullanılması, önerilen hibrit modelin performansının daha geniş ve karmaşık bir veri ortamında değerlendirilmesine önemli katkı sağlamıştır. METABRIC, geleneksel meme kanseri veri tabanlarına kıyasla çok daha büyük örneklem hacmine ve daha fazla klinik ile genomik değişkene sahiptir. Bu özellikleri sayesinde model, farklı hasta profilleri ve çok boyutlu veri yapıları üzerinde test edilmiş; böylece genelleme yeteneği daha güçlü bir şekilde doğrulanmıştır. Sonuçlar, hibrit mimarinin gerçek dünya koşullarına benzer karmaşık veri yapılarıyla etkili bir şekilde başa çıkabildiğini ve klinik tahmin süreçlerinde güvenilir bir yaklaşım olduğunu göstermektedir.

ACKNOWLEDGMENTS

First and foremost, I sincerely thank Almighty God for granting me the strength and patience to complete this dissertation.

I would like to express my deep gratitude to my supervisor, Prof. Dr. M. Fatih AKAY, for his valuable guidance and continuous support.

My thanks also go to the examination committee members: Prof. Dr. Mehmet Uğraş CUMA, Assoc. Prof. Dr. Serkan KARTAL, Prof. Dr. İbrahim Taner OKUMUŞ, Assoc. Prof. Dr. Fatih KILIÇ, for their constructive comments and the time they dedicated to reviewing my work.

I am grateful to the faculty of the Department of Computer Engineering at Çukurova University and to the Institute of Natural and Applied Sciences for their support throughout my studies.

I lovingly remember my late father, whose memory has remained a source of strength and inspiration. I dedicate this work to him.

My heartfelt thanks to my dear friends and colleagues, especially DM. Ala and Dr. FS. M, for their constant support and motivation along this journey.

My deepest appreciation goes to my family for their patience, love, and encouragement. I am especially thankful to my husband, Yahya ALSAEDI, for his unwavering support.

I also thank my children Dr. Nabaa, Jamal, Murtadha, and Jumana whose love and smiles have been a continuous source of motivation.

Finally, I extend my sincere thanks to my mother, siblings, and extended family for their prayers and continuous encouragement.

LIST OF TABLES

LIST OF FIGURES

SYMBOLS AND ABBREVIATIONS

AB : AdaBoost

ABC : Artificial Bee Colony

ACO : Ant Colony Optimization

ACSC : Australian Cyber Security Centre

ADMM : Alternating Direction Method of Multipliers

AE : Auto-Encoder

AI : Artificial Intelligence

AN : Artificial Neural Networks

AUC : Area Under the Curve

AWRF : ANN Weighted Random Forest

BN : Bayesian Network

CNN-VGG : Convolutional Neural Network using VGG

D.T : Decision Tree

DCIS : Ductal Carcinoma in Situ (Malignant)

Deep MLP : Deep Multilayer Perceptron

DT : Decision Tree

ENN : Ensemble Neural Network

GB : Gradient Boosting

GNB : Gaussian Naive Bayes

IDC : Invasive Ductal Carcinoma (Benign)

ILC : Invasive Lobular Carcinoma

KNN : K-Nearest Neighbors

LR : Logistic Regression

LSTM : Long Short-Term Memory Networks

METABRIC : Molecular Taxonomy of Breast Cancer International Consortium Dataset

ML : Machine Learning

MLP : Multilayer Perceptron

NB : Naive Bayes

NCC : Nearest Cluster Classifier

PPV : Positive Predictive Value

RF : Random Forest

ROC : Receiver Operating Characteristic Curve

SGD : Stochastic Gradient Descent

SVM : Support Vector Machine

SVR : Support Vector Regression

TNR : True Negative Rate

TPR : True Positive Rate

VC : Voting Classifier

WBCD : Wisconsin Breast Cancer Dataset

1. INTRODUCTION

Breast cancer is responsible for most deaths due to cancer in women all over the world. Therefore, it is very important to detect breast cancer at an early stage and classify it accurately for effective treatment and better patient outcomes. Although powerful diagnostic strategies (eg, mammography, biopsy, medical exams) exist, there are barriers [1-3]. Due to the slow, invasive, and error-prone nature of these techniques, diagnosis and treatment may be delayed [4, 5]. The standard types of diagnosis can be improved using machine learning and artificial intelligence [6-8]. By using ML algorithms that are capable of analyzing huge volumes of clinical records and recognizing different patterns, and offering accurate classifications. It can aid in the early detection of breast cancer [9-11]. Among various machine learning (ML) techniques, neural networks, decision trees, and autoencoders have shown proven potential in medical diagnostics [12-15]. The purpose of this is to deepen the investigation into integrating the strengths of these strategies together into a strong framework for breast cancer classification, as already achieved in this work.

The research uses the dataset called the Wisconsin Breast Cancer Dataset. This dataset was created by Dr. William H. Wahlberg, a Department of Medicine physician from the University of Wisconsin Hospital in Madison, Wisconsin, USA [15-17]. The new, complex algorithm was built from Deep MLPs, Feature-Fused Autoencoder, and Weight-Tuned Cross-Validated Decision Tree, trained and tested on these subsets. The performance parameters are as per accuracy, recall, F1 score, and precision. The complex algorithm optimized hidden units, dropout, and batch size, which affect performance parameters.

The chapter also offers a thorough and integrated overview of the medical and technical considerations of breast cancer, taking a medical look at breast cancer, which gives an overview of the disease and its global extent. Later, it discusses all its types, the symptoms most associated with it, as well as the treatment used. When diseases are detected early, the chances of recovery increase. Thus, this is highly essential.

Cancer refers to diseases that cause body cells to mutate or change in such a way that the cells replicate uncontrollably. Tumors and cancers can occur in any part of the body [18, 19]. Breast cancer is one of the types of abnormal growths, and the cancer cells mostly grow in the lobules or mammary glands in the breast [20, 21].

Breast cancer is classified as one of the most dangerous cancers because it does not usually exhibit symptoms or signs during the early development of the tumor [6, 22, 23]. Thus, it is vital to adopt effective early detection methods. Most women get a physical exam for a solid lump in the breast or swelling of the nodes in the armpit area [24-26].

Usually, a clinical examination will help your physician diagnose breast cancer, but most of the time, a patient finds out about it after she notices a lump. A mammogram is also used, which

is a machine that squeezes the breast tightly to see lumps underground [27-29]. A mammogram may highlight tumors that are most likely benign (non-cancerous), but we may perform a needle biopsy if diagnosis is suspected [28, 30, 31].

The most common type of breast cancer is invasive, defined as cancer cells breaking out of ducts or glands and invading surrounding breast tissue [32, 33].

Breast cancer is a broad term that covers several molecular types of breast cancer (4) and histological types (more than 21) [34, 35]. The risks, rates of recurrence, and clinical outcomes of these types are very different. The classification of histological types occurs based on their size, shape, and arrangement [36]. On the other hand, molecular types refer to the gene expression of the tumor and its biological markers, including estrogen receptor (ER), progesterone receptor (PR), and HER2 receptor [37, 38].

Cancer treatment strategies depend on the type of cancer and stage, age, and preference of the patient. The primary aim of the treatment is the removal of all cancer cells from the human body [39]. If surgery is required, a total mastectomy is the removal of an entire breast, or in some cases like a partial mastectomy where the tumor and a suitable margin of surrounding healthy tissue is removed [40]. Surgery is typically followed by radiation therapy to eliminate any remaining cells and reduce the risk of recurrence [41].

Besides local treatment, systemic treatment is also done. Systemic treatment uses medication that travels in the bloodstream and affects the whole body [42]. Although effective, these treatments can cause side effects on other tissues and organs. Even though effective, these treatments can affect other tissues and organs [43]. One treatment is chemotherapy, and, in general, they are used for metastatic cases and in young women. One can choose from several treatments for breast cancer according to the prescription of the doctor. Hormonal therapy, targeted therapy, and immunotherapy are among them [44].

In 2019, there were about 268,600 new cases of breast cancer and 41,760 deaths in the USA, according to the American Cancer Society [45, 46]. This shows that the cases of disease are increasing year by year and taking a toll on public health.

About twenty percent to thirty percent of people diagnosed with breast cancer experience recurrence. This is relatively high compared to other types of cancers. This means scientists need to keep researching how to treat cancer better [46, 47].

To develop better treatment protocols, it is important to review the medical records of previous patients and evaluate treatment outcomes.

There are several software tools available for extracting and analysing medical information from clinical documents. cTAKES is a programme that assesses the clinical notes of patients to extract relevant terms suggesting the recurrence of breast cancer, which can help monitor patients at risk of recurrence [48].

A breast cancer recurrence occurs when the cancer returns after the original lesion has been fully treated and believed to be cured. Remaining cancer cells that were not eliminated during the first course of treatment frequently cause recurrence [49]. In order to create treatment plans that lower recurrence rates and increase the likelihood of disease-free survival, it is essential to understand the causes of recurrence and the mechanisms by which it is detected.

1.1. Types of Breast Cancer

There are different types of breast cancer based on where the cancer cells start and how they spread. The most important groups are:

- Benign Tumors: These are growths that are not cancerous and do not spread to other parts of the body. Benign tumors are not life-threatening, but they may need to be watched or surgically removed if they cause pain[19, 50, 51].
- Malignant Tumors: These are cancerous and can spread to other parts of the body and invade nearby tissues. Malignant tumors necessitate immediate and vigorous intervention [52, 53].

Further classification includes:

- Ductal Carcinoma in Situ (DCIS): A non-invasive cancer in which abnormal cells are contained within the milk ducts [54, 55].
- Invasive Ductal Carcinoma (IDC): The most common type, where cancer cells spread beyond the ducts into the surrounding breast tissue [54, 56].
- Invasive Lobular Carcinoma (ILC): Cancer that begins in the lobules and invades surrounding tissues [57-59].

1.2. Symptoms and Risk Factors

Common symptoms of breast cancer include [60, 61]:

- A lump in the breast or underarm
- A change in the breast's size, shape, or look
- An unusual discharge from the nipple
- Dimpling or redness on the breast skin

Major risk factors include:

- Family history and genetic mutations (e.g., BRCA1, BRCA2)
- Age and gender
- Hormonal imbalances
- Obesity and lifestyle-related factors

1.3. Importance of Early Detection

Early detection of breast cancer greatly increases the likelihood of effective treatment and survival [62]. In clinical settings, they use mammography, ultrasound, and biopsy. But these old-fashioned ways of diagnosing can take a long time, cost a lot of money, and lead to errors due to human error [63].

Researchers are exploring how to combine smart systems, such as machine learning and deep learning algorithms, to help doctors make faster, more accurate, and more consistent diagnostic decisions. AI-based systems can analyze complex patterns in medical datasets to detect cancerous traits earlier.

1.4. Significance and Objectives of the Proposed Study

The reviewed studies primarily employed traditional machine learning models, without integrating deep learning or feature fusion strategies. They often had imbalances in class-wise performance or lacked strength in key metrics such as accuracy, recall, and AUC. In response, the proposed method in this thesis introduces a hybrid framework that combines Deep Multi-Layer Perceptron (MLP), Feature-Fused Autoencoder Neural Networks, and a Weight-Tuned Cross-Validated Decision Tree. This framework addresses the limitations of previous works by improving feature representation, enhancing generalization, and achieving consistently higher performance across all major evaluation metrics. The combination of deep learning and model optimization techniques enables more accurate and reliable breast cancer diagnosis using the WDBC dataset.

This study aims to develop an efficient machine learning (ML) framework that combines a deep multilayer perceptron (Deep MLP) for function extraction, a feature-fused autoencoder for dimensionality reduction, and a weight-tuned cross-validated decision tree for final classification. Each element of the developed framework plays a crucial role in enhancing overall performance. The Deep MLP is a sort of synthetic neural network that includes more than one layer of neurons, including input, hidden, and output layers. It is, in particular, powerful in extracting complex features from entered data, which might be crucial for accurate classification. By leveraging its deep knowledge of techniques, the Deep MLP can identify challenging patterns within the data that may be indicative of malignant or benign breast cancer. An autoencoder is an unmonitored neural network designed to study the encoding of entered data. This study introduces a Feature-Fused

Autoencoder that not only reduces the dimensionality of the input information but also enhances the representation, thereby gaining insight into the method. The fusion mechanism in the autoencoder integrates characteristic extraction and dimensionality reduction, resulting in a more compact and informative representation of the facts. This step is crucial for reducing computational complexity and enhancing the efficiency of the subsequent classification level. Decision trees are extensively used for classification duties due to their simplicity and interpretability. However, their overall performance may be substantially enhanced through weight tuning and cross-validation. The Weight-Tuned Cross-Validated Decision Tree in our framework adjusts the weights of various training sets to handle class imbalance. It employs cross-validation to ensure a robust and unbiased evaluation of the model. This method ensures that the final category version is accurate and generalizable. To evaluate the effectiveness of the proposed framework, we employ k-fold cross-validation. This robust technique divides the dataset into k subsets, allowing for multiple rounds of training and testing. This method provides a comprehensive assessment of the version's overall performance across distinct record splits, thereby reducing the likelihood of overfitting and ensuring the reliability of the effects. Performance metrics, including accuracy, precision, F1-rating, and Area Under the Curve (AUC), are used to quantify the version's effectiveness in classifying breast cancer.

1.5. The principal contributions of the thesis

The principal contributions of this study encompass:

1. Improvement of a singular ML framework that integrates Deep MLP, Feature-Fused Autoencoder, and Weight-Tuned Cross-Validated Decision Tree for breast cancer classification;
2. A comprehensive assessment of the proposed framework, the usage of the Wisconsin breast cancer dataset, demonstrating advanced performance
3. A deep analysis of the novel's complicated algorithms' overall performance using more than one metric, like accuracy, recall, F1-score, and precision.
4. Study the parameters that affect the performance parameters that can be adjusted in the complicated algorithm, like hidden units, dropout, and batch size.
5. Study the relationships between the influencing parameters and determine the extent of their impact on performance parameters.
6. By integrating current deep MLP techniques into a cohesive and practical framework, this study aims to contribute to ongoing efforts to improve breast cancer prognosis and, ultimately, enhance patient care and outcomes.

This thesis consists of five chapters. Chapter 1 introduces the research problem, the types of breast cancer, its symptoms, risk factors, the importance of early detection, the objectives of the study, its significance, and the main contributions. Chapter 2 provides a comprehensive review of the available literature on breast cancer classification using machine learning techniques. Chapter 3 reviews the basic theories underlying traditional machine learning methods and then explains the hybrid method proposed in this thesis. It also details the sub-algorithms it comprises. In addition, the chapter addresses performance benchmarks and begins by describing the steps for applying the hybrid algorithm to the selected breast cancer database. Chapter 4 provides an overview of the results across various hyperparameters, along with necessary discussions and comparisons with other researchers' findings. Finally, Chapter 5 discussed the conclusions and future work.

2. PRELIMINARY WORK

This chapter provides a comprehensive review of the literature on breast cancer classification using machine learning techniques. It summarizes the most significant research, detailing the algorithms used, datasets utilized, evaluation metrics, and key findings. Particular attention is paid to comparative studies, hybrid and deep learning methods, and explainable AI techniques. The chapter also highlights limitations and research gaps identified in previous work, providing a clear and justified basis for the methodology proposed in subsequent chapters.

(M. M. Islam et al., 2020) [63] applied five traditional machine learning algorithms, SVM, KNN, RF, ANN, and LR, to breast cancer classification using the WDBC dataset. Although ANN achieved strong results, the study relied solely on conventional models and did not incorporate deep learning or feature-level enhancements.

(Nirdosh Kumar, 2020) [64] Conducted a comparison using LR, SVM, KNN, and Naïve Bayes. While SVM performed the best, the other models showed limited recall and accuracy, and the study did not explore advanced optimization or deep representation learning techniques.

(G. Battineni et al., 2020) [65] evaluated SVM, LR, and KNN models. Despite having acceptable precision, their models showed lower accuracy, recall, and AUC. Their approach also did not integrate deep learning or feature fusion methods.

The study of P. Karatza, et al 2021 [66], which uses RF, NN, and ENN in the diagnosis of breast cancer from the WDBC dataset . This study is compared with the proposed work.

(Jiande Wu, and Chindo Hicks 2021) [67]. This study evaluated four different classification models, including Support Vector Machines, K-nearest neighbor, Naïve Bayes, and Decision Tree, using features selected for Breast Cancer Type Classification Using Machine Learning. SVM achieved the highest Accuracy and Recall were 90% and 87% respectively, while the highest Specificity was 91% with DT.

(Vu Pham Thao Vy et al.2022) [68], This study aimed to develop a machine-learning classification model to differentiate ductal carcinoma in situ and minimally invasive breast cancer using clinical characteristics, mammography findings, ultrasound findings, and histopathology features. The model of this study showed that the five most important features were calcifications on mammograms, lymph node presence, microcalcifications on histopathology, the shape of the mass on ultrasound, and the orientation of the mass on ultrasound.

(Mohammed Amine Naji et al. 2021) [69] This study used SVM, RF, LR, DT, and KNN for breast cancer prediction and diagnosis. They studied accuracy, precision, sensitivity, and F-Measure. SVM achieved the highest accuracy ≈ 97.2% on the WBCD dataset

(A. I. Aishwarja et al., 2020-2021) [70] Implemented RF, KNN, SVM, and Naïve Bayes. SVM showed the best performance among them; however, the models overall had limitations in precision and recall. The study did not include modern neural architectures.

(S. Sakib et al., 2021-2022) [71] employed a wide range of classifiers, including SVM, DT, LR, RF, KNN, and NN. While RF achieved strong F1-scores, most models underperformed in recall and accuracy. The work focused solely on traditional machine learning without using deep learning.

(Bharti Thakur et al. 2022) [72] Machine learning classifiers, SVM, Decision Tree, Ada Boost, random forest, and ANN, were used in this paper for detecting breast cancer. SVM gives the accuracy of 86.9%; decision tree 85.3%, random forest 86.3%, Ada Boost 68.5%, and ANN gives 87.43% accuracy.

(M. Monirujjaman Khan et al., 2022) [73] applied RF, LR, DT, and KNN, reporting class-wise F1-scores for benign and malignant cases. Although LR showed better class-wise results, the overall approach lacked depth in learning capability and optimization.

(V. Nemade and V. Fegade K, 2023) [74] evaluated multiple algorithms, KNN, SVM, DT, Naïve Bayes, and LR, based on class-wise performance metrics. Despite DT and LR yielding balanced scores, inconsistencies were observed across classes, particularly for malignant cases.

(Khandaker Mohammad Mohi Uddin et al. 2023) [75]. In this research work, the Wisconsin Breast Cancer Dataset (WBCD) has been used as a training set from the UCI machine learning repository to compare the performance of the various machine learning techniques. Different kinds of machine learning classifiers such as support vector machine (SVM), Random Forest (RF), K-nearest neighbors(K-NN), Decision tree (DT), Naïve Bayes (NB), Logistic Regression (LR), AdaBoost (AB), Gradient Boosting (GB), Multi-layer perceptron (MLP), Nearest Cluster Classifier (NCC), and voting classifier (VC) have been used for comparing and analyzing breast cancer into benign and malignant tumors. Various matrices, such as error rate, Accuracy, Precision, F1-score, and recall, have been implemented to measure the model's performance. Each Algorithm's accuracy has been ascertained for finding the best suitable one. Based on the analysis, the result shows that the Voting classifier has the highest accuracy, which is 98.77%, with the lowest error rate.

(Dinesh & Kalyanasundaram, 2023) [76] compared SVM, KNN, Decision Tree, and Random Forest using UCI's breast cancer data. Random Forest outperformed others with ~95% accuracy, versus 92% (SVM), 90% (KNN), and 88% (Decision Tree).

(Aiman Fatima et al. 2023)[77] In this study, different algorithms are examined. SVC obtained an 86.36%, Decision Tree an 86.18%, and Random Forest an 86.00%. The deep learning model, more precisely, a neural network, outperformed these results with a highest 93% accuracy.

(Ruiming Zeng et al. 2024) [78] In this work, an efficient and robust deep learning framework called FastLeakyResNet-CIR, an improved ResNet architecture, is proposed for breast cancer detection and classification

(Patnala S. R. et al. 2024) [79] The objective of this investigation was to improve the diagnosis of breast cancer by combining two significant datasets: the Wisconsin Breast Cancer

Database and the DDSM Curated Breast Imaging Subset (CBIS-DDSM). A hybrid deep learning algorithm consisting of CNNs and stochastic gradients is used for the detection of breast cancer, utilizing an accuracy is 96% precision of 95%, a Recall is 95%, and an AUC-ROC value of 0.96.

(Ghosh & Chatterjee, 2024, preprint) [80] evaluated ten ML algorithms (including SVM, XGBoost, CNN, RNN, Random Forest, and Logistic Regression) on UCI breast cancer data. SVM emerged best, achieving 98.25% accuracy, outperforming even CNN and XGBoost.

(Badar Almarri et al. 2024) [81]. Using machine learning techniques, including KNN, D.T., R.F., SVR, and Gaussian Naive Bayes (GNB), this research aims to create a breast cancer detection model. This research deduces that the Grid Search Algorithm outperforms all the other by achieving 92.6383% accuracy.

(Sharma et al., 2025)[82] compared Random Forest, Neural Networks, Logistic Regression, and tuned models using Randomized Search CV across WBCD and Coimbra datasets. Random Forest demonstrated superior accuracy and generalizability.

In a study by M. Güler et al. (2025) [83], DenseNet201, a deep convolutional neural network, was employed for breast cancer classification. The model achieved an accuracy of 89.4%, precision of 88.2%, recall of 84.1%, F1-score of 86.1%, and an AUC of 95.8%, underscoring the efficacy of deep learning in complex medical image analysis.

(H. Rafiepoor et al., 2025)[84] compared various ML models, including Random Forest, Support Vector Machine, and Logistic Regression, with the traditional Gail risk assessment model for breast cancer risk prediction. The ML models outperformed the Gail model in recall and F1-score, highlighting the potential of AI-based approaches in early risk prediction.

(T. Arravalli et al., 2025) [85] utilized Random Forest classifiers combined with explainable AI techniques for breast cancer detection. The model achieved an F1-score of 84%, and the integration of explainable AI methods provided transparency and interpretability to the model's predictions.

(Umesh Kumar Lilhore et al. 2025) [86]. The study named “Hybrid convolutional neural network and bi-LSTM model with EfficientNet-B0 for high-accuracy breast cancer detection and classification” used CNN + Bi-LSTM + EfficientNet-B0 and achieved an accuracy score of 99.2%

Table 2.1 Literature review

Authors and Year	Objective	Methods	Evaluation Metrics	Obtained Results
M. M. Islam et al., 2020	Breast cancer classification	SVM, KNN, RF, ANN, LR	Accuracy, F1-score	ANN achieved strong results (~98%); no deep learning or feature enhancement
Nirdosh Kumar, 2020	Comparison of ML models	LR, SVM, KNN, Naïve Bayes	Accuracy, Recall	SVM best (~97%); limited recall in others
G. Battineni et al., 2020	Model evaluation for diagnosis	SVM, LR, KNN	Precision, Recall, AUC	Acceptable precision; lower recall & AUC
M. Vakili et al., 2020	Comparison of ML and DL methods	11 ML and DL algorithms	Accuracy, F1-score	RF highest F1-score; ensemble models effective
Wu, J., & Hicks, C. (2021)	To classify triple-negative breast cancer (TNBC) and non-TNBC subtypes using RNA-Seq gene expression data from TCGA.	SVM, KNN, Naïve Bayes, and Decision Tree	Accuracy, Precision, Recall, F1-score	SVM achieved the highest performance with approximately **90% accuracy**. Other models showed lower recall and precision
P. Karatza et al., 202	Breast cancer diagnosis improvement	RF, NN, ENN	Recall, AUC, Precision	ENN outperformed RF and NN with improved metrics

Table 2.1. Literature review (continued)

Mohammed Amine Naji et al. 2021	Breast Cancer Prediction and Diagnosis	SVM, RF, LR, DT, and KNN	Accuracy, Precision, Sensitivity, and F-Measure	SVM achieved highest accuracy ≈ 97.2% on WBCD dataset
A. I. Aishwarja et al., 2021	Model comparison on WDBC data	RF, KNN, SVM, Naïve Bayes	Accuracy, Recall, Precision	SVM best; limited precision and recall overall
S. Sakib et al., 2022	Evaluation of ML classifiers	SVM, DT, LR, RF, KNN, NN	F1-score, Recall	RF achieved highest F1 (~94%)
Bharti Thakur et al., 2022	Detection using ML classifiers	SVM, DT, AdaBoost, RF, ANN	Accuracy	ANN (87.43%), SVM (86.9%), RF (86.3%)
M. Monirujjaman Khan et al., 2022	Class-wise breast cancer classification	RF, LR, DT, KNN	F1-score	LR showed better class-wise results
V. Nemade & V. Fegade, 2023	Comparison of ML algorithms	KNN, SVM, DT, Naïve Bayes, LR	Precision, Recall	DT and LR yielded balanced scores
Khandaker Mohammad Mohi Uddin et al. 2023	Wisconsin Breast Cancer Dataset (WBCD) from UCI repository	NB, LR, AdaBoost (AB), Gradient Boosting (GB), Multi-Layer Perceptron (MLP), Nearest Cluster Classifier (NCC), and Voting Classifier (VC).	Error rate, Accuracy, Precision, F1-score, Recall	GB highest accuracy (97.9%)
Dinesh & Kalyanasundaram , 2023	Algorithm performance comparison	SVM, KNN, DT, RF	Accuracy	RF (95%) > SVM (92%) > KNN (90%)

Table 2.1. Literature review (continued)

Aiman Fatima et al., 2023	ML vs DL for classification	SVC, DT, RF, Neural Network	Accuracy	NN achieved 93% accuracy; DL better overall
Ruiming Zeng et al. 2024	Optimized CNN architectures	FastLeakyResNet-CIR ResNet50 InceptionV3 ResNet18 VGG16	Accuracy	FastLeakyRes Net-CIR achieved highest accuracy 98.94%
Patnala S. R. et al., 2024	Hybrid deep learning model	CNN + SGD	Accuracy, Precision, Recall, AUC	Accuracy 96%, Recall 95%, AUC 0.96
Ghosh & Chatterjee, 2024	Comprehensive model benchmark	SVM, XGBoost, CNN, RNN, RF, LR	Accuracy	SVM achieved 98.25%; outperformed CNN & XGBoost
Badar Almarri et al., 2024	Detection model optimization	KNN, DT, RF, SVR, GNB	Accuracy	Grid Search achieved 92.64% accuracy
Sharma et al., 2025	Cross-dataset generalization	RF, NN, LR, Tuned models	Accuracy, Generalizability	RF superior; best overall performance
M. Güler et al., 2025	Deep CNN for classification	DenseNet201	Accuracy, Precision, Recall, AUC	Acc 89.4%, Prec 88.2%, Rec 84.1%, AUC 95.8%
H. Rafiepoor et al., 2025	Risk prediction using ML	RF, SVM, LR vs Gail model	Recall, F1-score	ML outperformed Gail model
T. Arravalli et al., 2025	Explainable AI for detection	Random Forest + XAI	F1-score	RF with XAI achieved F1 = 84%
Umesh Kumar Lilhore et al., 2025	Hybrid DL model for detection	CNN + Bi-LSTM + EfficientNet-B0	Accuracy	Achieved 99.2% accuracy

Previous studies demonstrate the effectiveness of some traditional machine learning algorithms, such as SVM, RF, KNN, and LR, in classifying breast cancer. However, they also reveal several shortcomings. Most previous research has focused solely on traditional models without leveraging modern deep learning methods, feature fusion techniques, or advanced optimization techniques. While some models achieved acceptable accuracy, they suffered from low

recall, poor generalization across different databases, and fluctuating performance across classes, especially in malignant cases. Furthermore, few studies have examined hybrid approaches that combine the advantages of more than one algorithm, and explainable AI techniques have not received sufficient attention. The proposed work is a hybrid approach that combines three complementary algorithms: the deep multilayer neural network (Deep MLP), the feature-fused autoencoder, and the weight-tuned cross-validation decision tree. Through this integration, the proposed system aims to increase classification accuracy, improve recall, and achieve higher and more reliable generalization in breast cancer diagnosis.

Previous studies demonstrate the effectiveness of some traditional machine learning algorithms such as SVM, RF, KNN, and LR in breast cancer classification, but they also reveal several weaknesses. Most previous research has focused solely on traditional models without leveraging modern deep learning methods, feature fusion techniques, or advanced optimization techniques. Some models achieved acceptable accuracy but suffered from low recall, poor generalization across different databases, and fluctuating performance across categories, especially in malignant cases. Furthermore, few studies have addressed hybrid approaches that combine the advantages of more than one algorithm.

Given these challenges, this work proposes a hybrid approach that combines three complementary algorithms: a deep multilayer neural network (Deep MLP), a feature-fused autoencoder, and a weight-tuned cross-validation decision tree. Through this integration, the proposed system aims to increase classification accuracy, improve recall, and achieve higher and more reliable generalization in breast cancer diagnosis.

3. MATERIAL AND METHOD

With technical advancement, harmful diseases that threaten human survival are increasing at a similar rate. For women, breast cancer is considered the second most deadly disease, and it can be curable if it is detected at an early stage. In day-to-day health applications, the earlier diagnosis of breast cancer based on an automated process is expected [87]. Breast cancer manual diagnosis is time-consuming and tedious; thus, automatic diagnosis is essential. The recent healthcare system is effective; however, they are prone to faults. Artificial intelligence based on medical image classification is emerging as an effective tool that assists doctors by classifying medical images in various characteristics, resulting in early treatment and diagnosis. Machine learning and deep learning are comprised of various algorithms that are capable of diagnosis that is more accurate at an early stage, reducing the mortality rates. Artificial intelligence approaches boost the development of a new model in medical image classification to reduce healthcare expenses due to faulty diagnosis [88]. Medical imaging refers to certain approaches that are used to evaluate the human body to diagnose, treat, or track abnormalities. Medical imaging researchers are classifying the size, characteristics, and location of the abnormalities as an efficient approach to extracting useful data from huge amounts of information. Various researchers intensively focus on the interpretation and production of medical images to identify the majority of abnormalities. Machine learning and deep learning have made essential progress in recent decades and play an important role in the medical field, such as medical image classification.

Significantly, medical imaging is considered an effective way to identify breast cancer with various modalities like PET, mammography, MRI, ultrasound, CT, and radiography, which are regularly used for the breast cancer screening process.

3.1. Breast Cancer Data Sources

The development of any machine learning-based diagnostic system relies heavily on the availability of quality datasets. In the context of breast cancer classification, numerous datasets have been collected and made publicly available to support research in automated diagnosis. These datasets include clinical features, demographic information, pathological data, and imaging results. For this study.

3.1.1. Wisconsin Diagnostic Breast Cancer (WDBC) dataset

The Wisconsin Diagnostic Breast Cancer (WDBC) dataset has been selected due to its popularity, accessibility, and suitability for classification tasks.

This dataset was downloaded from the UCI Machine Learning Repository, a well-known and reliable source of datasets used in AI research. A simplified version of this dataset includes 569

samples (examples) and 32 features (features), with each sample representing a fine-needle aspiration (FNA) sample taken from breast tissue.

Fine needle aspiration (FNA) is a non-invasive diagnostic procedure that uses a thin needle to extract cells from suspicious tissue for microscopic examination. The WBCD dataset contains samples classified as malignant (212 cases) and benign (357 cases) and is free of missing values for all features.

Each sample is represented as a 9-dimensional vector with values ranging from 1 (most normal) to 10 (most abnormal), based on the extracted nucleus features. The dataset includes 10 key numerical features calculated for each cell nucleus:

Description of cell nucleus properties in the WBCD dataset [89].

1. Radius: The arithmetic mean of the distances from the center of the cell nucleus to points on its periphery. It is used as a general indicator of nucleus size.
2. Texture: The standard deviation of grayscale values within the nucleus, reflecting local variation in illumination intensity and image composition.
3. Perimeter: The total length of the nucleus boundary, calculated based on the connected points that form the nucleus periphery.
4. Area: The number of pixels within the nucleus boundary, plus half the number of pixels on the periphery. It is used as a measure of nucleus size.
5. Smoothness: Measures local variations in radius lengths by calculating subtle differences between the distances from the center to the edges. Lower values indicate smoother, more uniform boundaries.
6. Compactness: Calculated using the equation: Perimeter^2 / area. It expresses the density or compactness of the nucleus's shape.
7. Concavity: Measures the intensity of the concave segments (inward-facing areas) around the nucleus' perimeter. Higher values indicate a more irregular and complex boundary.
8. Concave Points: The number of concave segments or gaps evident on the nucleus' perimeter.
9. Symmetry: Expresses the asymmetry of the nucleus's shape by measuring the differences in the lengths of lines drawn perpendicular to the nucleus's major axis in both directions to the nucleus's boundary.
10. Fractal Dimension: A measure of boundary complexity using the "coastline estimation" method. Higher values indicate more irregular and jagged borders, which often indicate a higher likelihood of cancer cells being present.

For each of these ten features, three statistical indicators are calculated:

- Mean
- Standard Error (SE)
- Worst Value (Average of the three highest recorded values)

This results in 30 features per sample. For example, feature 2 represents the mean radius, feature 12 represents the standard error of the radius, and feature 22 represents the worst radius value. Figure (3-1) explains the correlation matrix of the WDBC dataset**.**

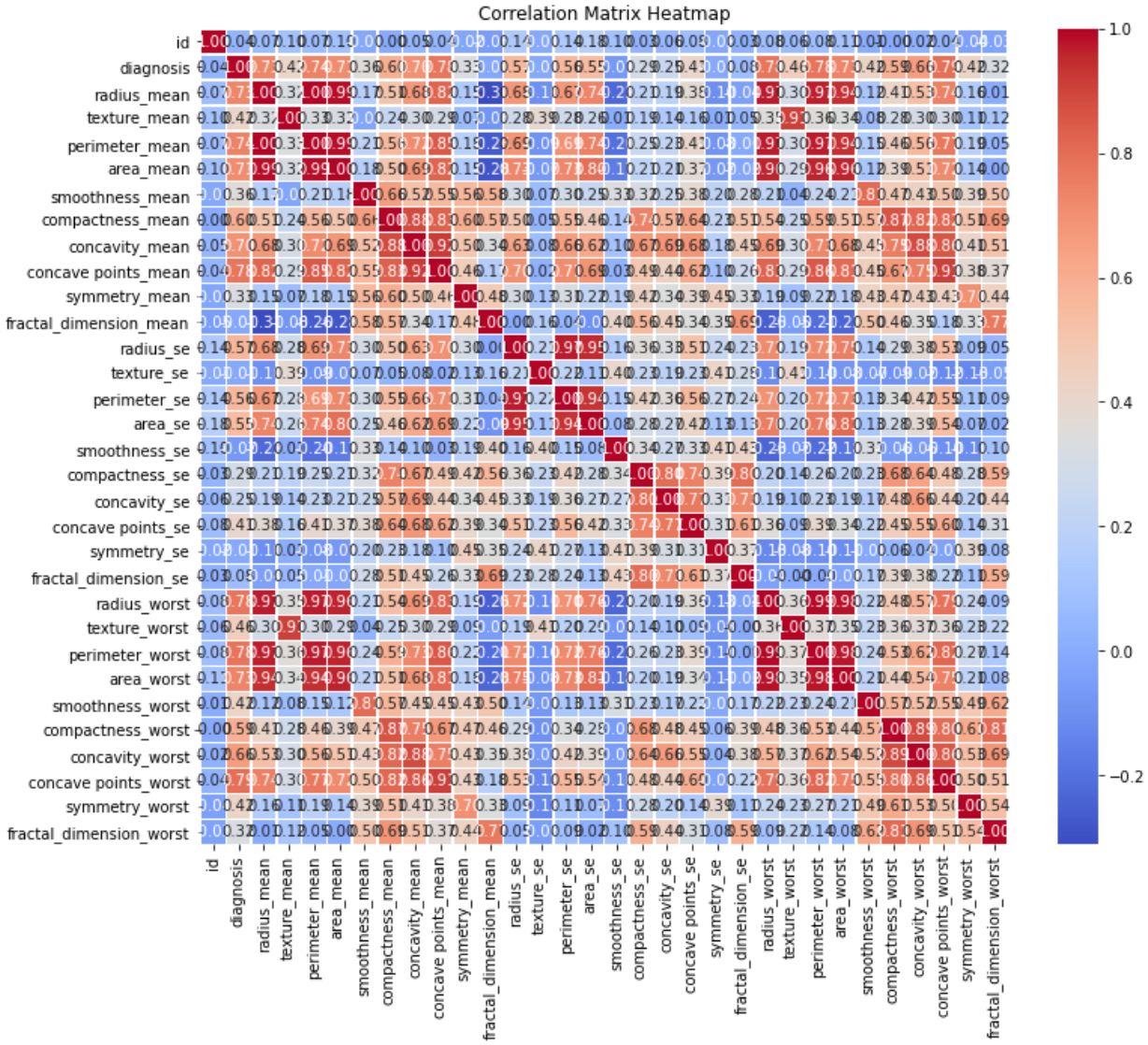


Figure 3.1. Correlation matrix of the WDBC dataset

3.1.2. Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset

The METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) dataset constitutes one of the most comprehensive and rigorously curated population-based cohorts in the field of breast cancer research. It comprises a total of 1,904 patients with extensive longitudinal follow-up, providing an exceptionally rich source of clinico-genomic information. The dataset encompasses detailed demographic characteristics, tumor histopathological features, hormone receptor profiles, therapeutic interventions, and meticulously documented survival outcomes. Moreover, METABRIC incorporates high-dimensional molecular data, including mRNA expression signatures and genome-wide copy number alterations spanning approximately 25,000 genes. The integration of clinical and genomic variables within a single dataset renders METABRIC particularly advantageous for the development of sophisticated predictive models in machine learning and deep learning. Its substantial cohort size, robust annotations, and prolonged follow-up duration significantly enhance the validity, reliability, and generalizability of analytical findings, thereby establishing METABRIC as a foundational and widely referenced resource in contemporary breast cancer researches [90, 91].

The METABRIC dataset is widely used in cancer research for:

- Survival prediction modelling
- Treatment outcome analysis
- Biomarker discovery
- Clinical decision support systems

Feature description of Metabric as follows [91, 92]:

1. Age at Diagnosis: The patient's age when breast cancer was first diagnosed.
2. Type of Breast Surgery: The type of surgery performed, such as mastectomy or breast-conserving surgery.
3. Cancer Type: High-level classification of the cancer (breast cancer or breast sarcoma).
4. Detailed Cancer Type: Specific histopathological subtype such as ductal, lobular, mucinous, metaplastic, etc.
5. Cellularity: The amount of residual tumor cells after chemotherapy.
6. Chemotherapy: Indicates whether chemotherapy was administered.
7. PAM50 + Claudin-Low Subtype: Molecular subtype derived from gene-expression profiling.
8. Cohort: The study group or batch to which the patient belongs.

9. Estrogen Receptor Status (IHC): ER expression measured by immunohistochemistry.
10. Estrogen Receptor Status: Indicates ER-positive or ER-negative classification.
11. Histologic Grade: The microscopic aggressiveness of the tumor (grades 1–3).
12. HER2 Status (SNP6): HER2 amplification measured using SNP6.
13. HER2 Status: Indicates HER2-positive or HER2-negative cancer.
14. Other Histologic Subtype: Additional tissue-based subtype classifications.
15. Hormone Therapy: Whether hormonal treatment was given.
16. Menopausal State: Patient's menopausal status (pre or post)
17. Integrative Cluster: Molecular classification based on multi-omic integration.
18. Tumor Laterality: Indicates whether the tumor is in the right or left breast.
19. Positive Lymph Nodes: Number of lymph nodes affected by the tumor.
20. Mutation Count: Total number of somatic mutations identified.
21. Nottingham Prognostic Index: Composite score predicting prognosis.
22. OncoTree Code: Standardized oncology diagnostic code.
23. Overall Survival (Months): Duration of survival after diagnosis or treatment.
24. Overall Survival Status: Indicates whether the patient is alive or deceased.
25. Progesterone Receptor Status: Indicates PR-positive or PR-negative tumor.
26. Radiotherapy: Whether radiation therapy was administered.
27. Three-Gene Classifier Subtype: Molecular subtype derived from ER, PR, and HER2.
28. Tumor Size: Size of the primary tumor.
29. Tumor Stage: Stage describing tumor spread and involvement.
30. Death from Cancer: Whether death was caused by breast cancer.

Figure 3.2. explains the correlation matrix of the METABRIC dataset**.**

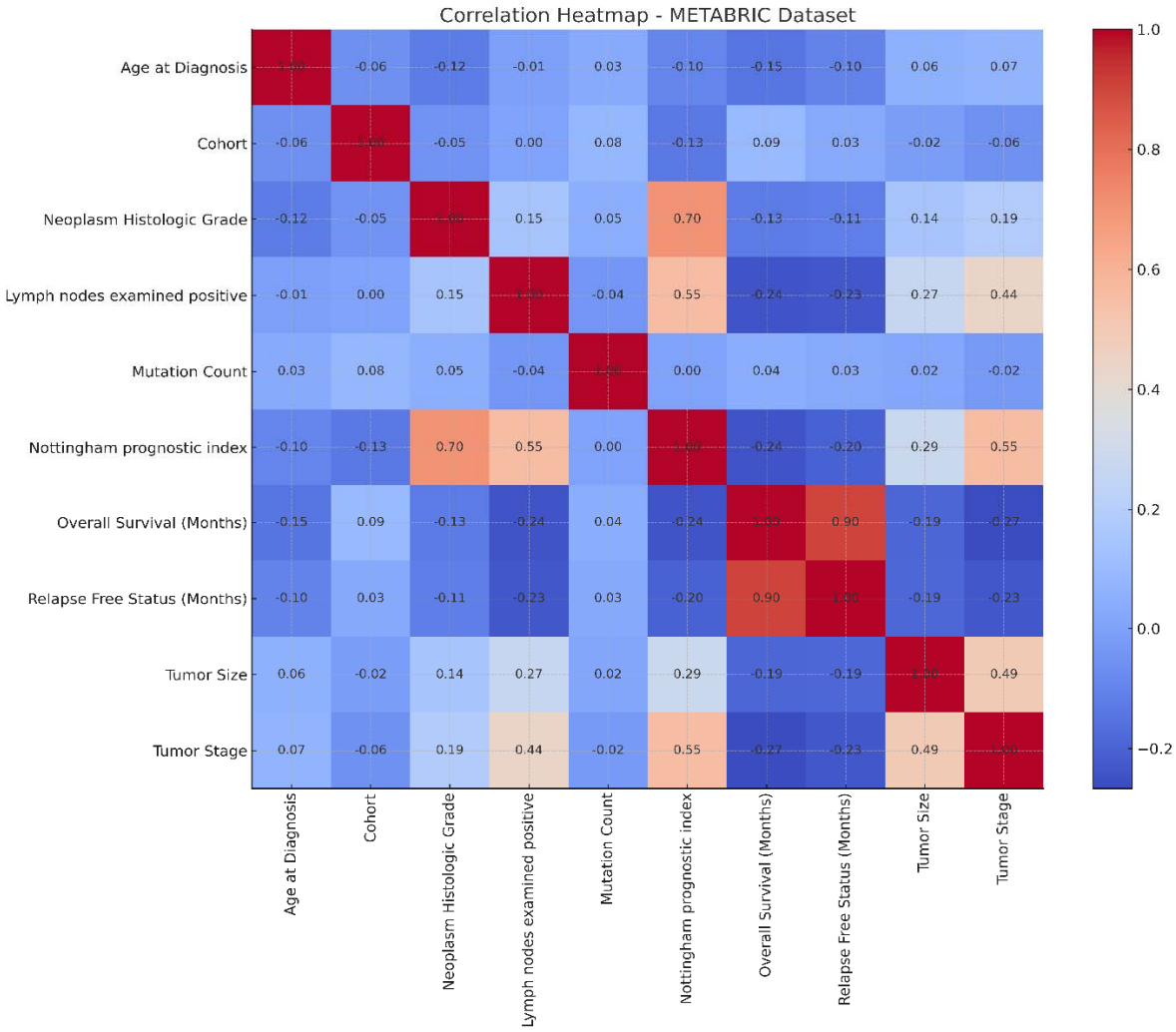


Figure 3.2. Correlation matrix of the METABRIC dataset

3.2. Diagnostic methods

3.2.1. Traditional methods

Many traditional machine learning methods are used to detect breast cancer. These methods rely on analysing features extracted from images or data of breast cells to classify them as benign or malignant. Here is a brief overview of the most important of these methods [93-96]:

A. Decision Tree:

This is an algorithm that builds a tree-like model, where data is progressively partitioned according to specific features. Each internal node represents a condition on a feature, and each branch represents the outcome of that condition. Ultimately, the leaves lead us to a final decision: is the sample malignant or benign? It is easy to interpret, but it can suffer from overgeneralization if not properly tuned.

B. Support Vector Machine (SVM):

It aims to find the optimal threshold between malignant and benign samples, ensuring that this threshold is as large as possible between the two classes. It is particularly effective in cases involving a large number of features, and its performance can be tailored using various types of kernels to suit the nature of the data.

C. Logistic Regression:

This is a statistical model used to estimate the probability that a sample belongs to a particular class (malignant or benign). Despite its simplicity, it is considered a powerful model for binary classification and is often used as a benchmark when evaluating the performance of more complex models.

D. K-Nearest Neighbors (KNN):

This model compares the new sample with its k nearest samples in the training set and determines the class based on the majority. That is, if most of the new sample's neighbors are malignant, the model classifies it as such. This method is simple but can become slow with large datasets.

E. Naive Bayes:

This is a classification model based on Bayes' theorem, assuming that each characteristic is independent of the others (a simplistic assumption that often fails to hold in practice). Despite its simplicity, it has demonstrated good performance in certain cases, particularly when the data is not complex.

F. Multi-Layer Perceptron (MLP):

This is a type of artificial neural network, consisting of multiple layers of nodes (neurons). It is used to learn complex patterns from data through a training process based on backpropagation. Although it is an older model, it remains in use in many medical applications.

3.2.2. Modified method (proposed method)

Traditional machine learning algorithms for breast cancer diagnosis face several significant challenges, including poor generalization ability when dealing with small or imbalanced datasets, inefficiencies in automatically extracting optimal features, reliance on manual input or manual feature selection, and difficulty representing nonlinear and complex relationships between different features.

To address these limitations, this work proposes a modified hybrid approach that integrates three powerful and complementary algorithms within a unified framework, specifically designed to

enhance the accuracy and reliability of breast cancer diagnosis. This modified method consists of a deep MLP, feature-fused autoencoder, and a weight-tuned cross-validation decision tree.

A. Deep Multilayer Perceptron (Deep MLP)

MLP refers to a feed-forward neural network (FNN) that passes data in a single direction through neural networks and their equivalent neurons, which are organized in multiple layers in a parallel manner. Initially, the first layer is the input layer, and the output layer is the last; the hidden layers are presented in the intermediate. When the FNNs are comprised of a single hidden layer, then it is referred to as an MLP.Figure3-3 explains the Deep MLP algorithm.

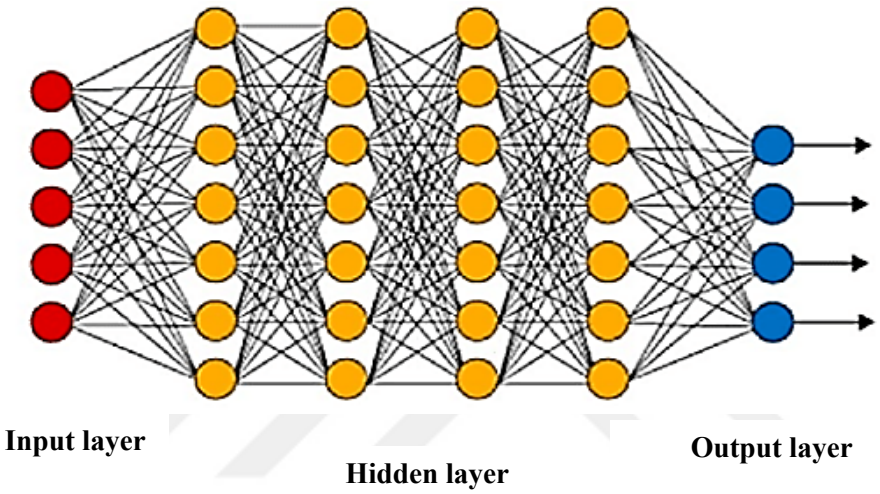


Figure 3.3. Deep MLP architecture[97]

The advantages considered are the ability of generalization and effective classification performance. Moreover, the systematic work is illustrated by variations between traditional machine learning methods and deep learning methods, as shown in Figures 3-4 and 3-5.

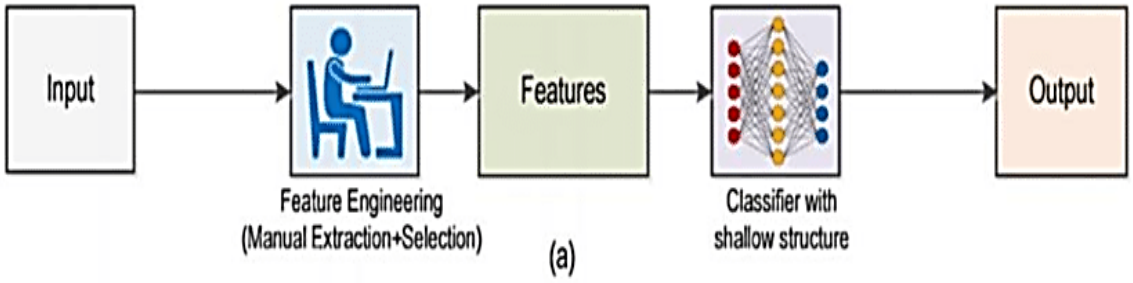


Figure 3.4.Traditional Machine Learning

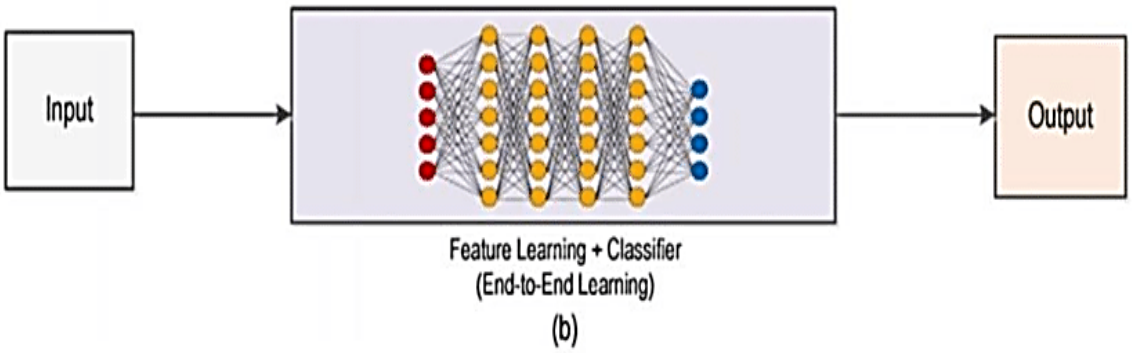


Figure 3.5. Structure of Deep Learning

Figure (3-3), shows (a) Shows that the traditional ML needs manual feature extraction. Whereas the modern DL models can automate every feature as well as the training process, thereby performing efficient learning. This exhibits the efficacy of the proposed DL over the traditional ML.

Mathematically, the deep MLP for Feature Extraction can be described as follows[98]:
Let the input data be:

$$\mathbf{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}\}, \mathbf{x}^{(i)} \in \mathbb{R}^d \quad (3\text{-}1)$$

Pass each input through a deep MLP:

$$\mathbf{h}^{(i)} = f_{\text{MLP}}(\mathbf{x}^{(i)}) = \sigma_L(\mathbf{W}_L \cdots \sigma_1(\mathbf{W}_1\mathbf{x}^{(i)} + \mathbf{b}_1) + \cdots + \mathbf{b}_L) \quad (3\text{-}2)$$

Where:

σ_j: activation function (e.g., ReLU)

$\mathbf{W}_j$, $\mathbf{b}_j$: weights and biases at layer j

$\mathbf{h}^{(i)} \in \mathbb{R}^{dd'}$: extracted feature vector

B. Feature Fused Auto Encoder Neural Network

A new structure of auto-encoder, namely, Feature Fused Auto Encoder Neural Network, is established based on the conventional auto-encoder symmetric property. Compared with the traditional auto-encoder, the new fused auto-encoder is generated by folding the traditional structure's right side to the left side. This feature fused auto-encoder contains Li1/2 hidden layers, similar to a traditional auto-encoder with L hidden layers. However, the variation is the computational cost among the two structures presented in the number of tuned weights, limited to half in the newly developed auto-encoder.

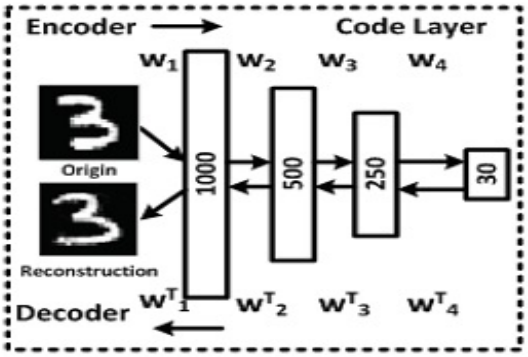


Figure 3.6.Feature Fused Auto Encoder Neural Network Architecture

Mathematically, the Feature-Fused Autoencoder can be described as follows: [99]

Let the extracted features be:

$$\mathbf{H} = \{\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \ldots, \mathbf{h}^{(N)}\} \tag{3-3}$$

The autoencoder encodes each feature vector into a lower-dimensional fused latent vector:

$$\mathbf{z}^{(i)} = f_{\text{enc}}(\mathbf{h}^{(i)}) \in \mathbb{R}^{d_z}, d_z < d' \tag{3-4}$$

Fusion Layer (e.g., attention or weighted average of multi-layer features):

$\mathbf{z}_f^{(i)} = \mathcal{F}(\mathbf{z}^{(i)})$, (e.g., self-attention, max-pooling, etc.)

Decoder (optional, for unsupervised training):

$$\hat{\mathbf{h}}^{(i)} = f_{\text{dec}}\left(\mathbf{z}_f^{(i)}\right) \tag{3-5}$$

Loss (if unsupervised reconstruction is used):

$$\mathcal{L}_{\text{AE}} = \frac{1}{N}\sum_{i=1}^{N} \left\|\mathbf{h}^{(i)} - \hat{\mathbf{h}}^{(i)}\right\|^2 \tag{3-6}$$

Now $\mathbf{z}_f^{(i)}$ are the dimensionality-reduced features.

C. Weight-Tuned Cross-Validated Decision Tree

A decision tree is regarded as a general classification of the divide-and-conquer approach. Based on this model, the data is represented in a tree model, the different characteristics are represented by internal nodes, and leaf nodes are expressed from data sample labels. The tree's

path is between the leaf and the root to find the corresponding dataset. The leaf node consists of the final grading outcome.

Mathematically, the Weight-Tuned Decision Tree can be described as follows: [98]

Train a decision tree classifier on reduced features. $\mathbf{z}_f^{(i)}$

Let $y^{(i)}$ be the label (e.g., benign/malignant)

The classifier is trained as:

$$\hat{y}^{(i)} = f_{\text{tree}}\left(\mathbf{z}_f^{(i)}; \theta\right) \tag{3-7}$$

Where: f_{tree} : is a decision tree, θ includes split criteria and sample weights tuned via cross-validation:

$$\theta = \arg\min_{\theta} \mathcal{L}_{\text{tree}} + \lambda \cdot \Omega(\theta) \tag{3-8}$$

$\mathcal{L}_{\text{tree}}$: classification loss (e.g., Gini impurity or entropy)

$\Omega(\theta)$: regularization (e.g., tree depth, number of leaves)

λ: regularization strength

The best θ is selected via cross-validation, optimizing performance metrics like accuracy, F1-score, or AUC.

To generate the proposed modified algorithm, the above three algorithms were combined, so the final equation of the combined algorithms is:

$$\mathbf{x}^{(i)} \xrightarrow{\text{Deep MLP}} \mathbf{h}^{(i)} \xrightarrow{\text{FFA}} \mathbf{z}_f^{(i)} \xrightarrow{\text{WT-CVDT}} \hat{y}^{(i)} \tag{3-9}$$

3.3. Performance Matrices

Firstly, the confusion matrix is one of the estimates used to evaluate the performance of the model:

3.3.1. Confusion Matrix

It is just a table that tells us which values were correctly predicted and which are not.

	Predicted Class Yes (+1)	Predicted Class No (-1)
Actual Class Yes (+1)	TP	FN
Actual Class No (-1)	FP	TN

Then the performance matrices that can be calculated are [100]

3.3.2. Sensitivity

Sensitivity/recall/true positive rate (TPR): It evaluates what percentage of patients are correctly identified to have a particular disease (+ ve instances) and is illustrated as:

$$S_{EN} = \frac{T_P}{T_P + F_N} \tag{3-10}$$

Where T_P is True Positives and F_N is False Negatives.

3.3.3. Specificity

Specificity/true negative rate (TNR): Determines what percentage of patients are correctly identified not to have a particular disease (−ve instances) and elucidated as:

$$S_P = \frac{T_N}{T_N + F_P} \tag{3-11}$$

where T_N is True Negatives, F_P is False Positives, and F_N is False Negatives.

3.3.4. Precision

Positive predicted value (PPV): Specifies what percentage of patients is actually relevant (+ ve) and described as:

$$P_R = \frac{T_P}{T_P + F_P} \tag{3-12}$$

where T_P is True Positives, and F_P is False Positives.

3.3.5. Accuracy

Accuracy: It is the ratio of correctly categorized samples to the total number of samples and is interpreted as:

$$AC_U = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \quad (3\text{-}13)$$

Where T_P is True Positives, T_N is True Negatives, F_P is False Positives, and F_N is False Negatives.

3.3.6. F1 Score

The F1 score is the harmonic mean of the precision and recall. It is also called the F Score or the F-measure.

In simple terms, the F1 score conveys the balance between precision and recall (sensitivity).

$$F1\ Score = \frac{2T_P}{2T_P + F_P + F_N} \quad (3\text{-}14)$$

3.4.The procedure steps of the proposed algorithm

The proposal proposes effective machine learning techniques to classify breast cancer from the Wisconsin breast cancer dataset effectively. Initially, the data is pre-processed to remove unwanted information. Further, the feature extraction of significant information or abnormalities is performed using the Deep Multi-Layer Perceptron (Deep MLP) approach. Moreover, dimensionality reduction is performed to avoid unnecessary or irrelevant data using the newly developed Feature Fused Auto Encoder Neural Network. Finally, the breast cancer classification is performed by using the Weight-Tuned Cross-Validated Decision Tree approach. The overall flow is shown in the following Figure 3-7.

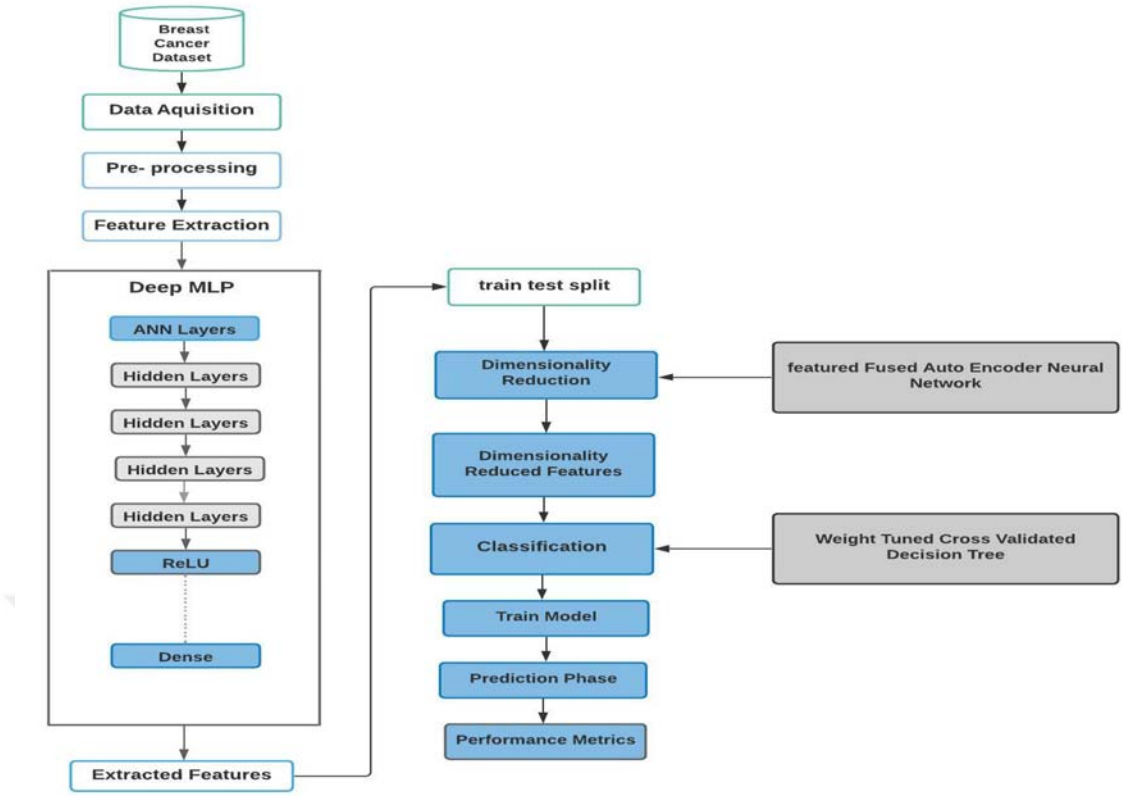


Figure 3.7. Flow chart diagram of the modified algorithm

3.4.1. Import the libraries

The libraries utilized in this study include Pandas, NumPy, Keras, and modules from Scikit-learn, specifically preprocessing and MinMaxScaler.

3.4.2. Exploratory data analysis

In this step, the DATASET has been read through the command

df = pd. read_csv("breast_cancer.csv")

3.4.3. Splitting into X and Y

In this step, the Wisconsin dataset (569, 32) was split into x and y, where x represents the data frame consisting of ID, diagnosis, and 32 other rows, while y represents diagnosis only, as shown in Table 3-1.

Table 3.1. x and y split data

x – DataFrame

Index	radius_ mean	texture_ mean	perimeter _mean	area_mea n	smooth ness_me an	compactnes s_mean	concavity _mean	concave points_ mean
1	0.1799	0.1038	0.1228	0.1001	0.1184	0.2776	0.3001	0.1471
2	0.2057	0.1777	0.1329	0.1326	0.0847	0.07864	0.0869	0.0701
3	0.1969	0.2125	0.1300	0.1203	0.1096	0.1599	0.1974	0.1279
4	0.1142	0.2038	0.7758	0.3861	0.1425	0.2839	0.2414	0.1052
5	0.2029	0.1434	0.1351	0.1297	0.1003	0.1328	0.1980	0.1043
6	0.1245	0.1570	0.8257	0.4771	0.1278	0.17	0.1578	0.0808
7	0.1825	0.1998	0.1196	0.1040	0.0946	0.109	0.1127	0.0740
8	0.1371	0.2083	0.9020	0.5779	0.1189	0.1645	0.0936	0.0598
9	0.1300	0.2182	0.8750	0.5198	0.1273	0.1932	0.1859	0.0935
10	0.1246	0.2404	0.8397	0.4759	0.1186	0.2396	0.2273	0.0854

y – Series

Index	diagnosis
1	M
2	M
3	M
4	M
5	M
6	M
7	M
8	M
9	M
10	M

3.4.4. Data normalization for the feature extraction

In this step, y, normalized as shown in table 3-3.

Table 3.2. The normalization of the y data

	0	**1**
0	0	1
1	0	1
2	0	1
3	0	1
4	0	1
5	0	1
6	0	1
7	0	1
8	0	1

3.4.5. Deep MLP for feature extraction

MLP is referring to a feed-forward neural network (FNN) that passes the data in an individual direction through a neural network and its equivalent neurons, which are organized in multiple layers in a parallel way. Initially, the first layer is the input layer, and the output layer is the last; the hidden layers are presented in the intermediate. When the FNNs comprise a single hidden layer, it is referred to as ML.

Deep MLP is creating multiple hidden layers in a neural network. Here, the depth refers to the number of hidden layers.

There are ten steps to train a Deep MLP as follows:

1. **Pre-process Data:** This is a very important step, and more attention should be paid here. Pre-processing helps us to understand the data better, and any irrelevant data can be removed, or the most important fields can be identified.
2. **Weights:** Weights need to be initialized before proceeding to the next step.
3. **Activation function:** ReLU is the most commonly used activation function now.
4. **Batch Normalization** is a process of adding a layer deep in the network to normalize the data. It enables faster convergence.
5. **Optimizer:** Most commonly used Optimizer is Adam (Adaptive Model Estimation)
6. **Hyperparameters:** Architecture — number of layers, number of neurons; Dropout rate; Optimizer parameters like mean/ variance at time (t).
7. **Loss Function:** Determine the log function. The commonly used log functions are log-loss, squared loss, etc.

8. **Gradient:** It is a good practice to monitor gradients and updates for each epoch, layer, and weight. Clipping is a common method for gradient monitoring.
9. **Plot:** Visualize train loss against epochs. The training loss should be approximately close to the testing loss for a perfect model.
10. **Over-fitting:** Take extra care to avoid overfitting. The difference between train and test loss should be small.

The advantages considered are the ability to generalize and effective classification performance. Table (3-3) represents the (x_feat) array.

Table 3.3. The (x_feat) array

	94	**95**	**96**	**97**	**98**	**99**
561	12.8344	-5.3651	20.4201	-17.4405	-15.8429	-4.8733
562	23.4714	-13.6170	38.1241	-37.8403	-29.7806	-12.9156
563	42.0273	-29.8629	69.3633	-76.2040	-54.4759	-28.8447
564	45.6611	-33.6410	75.5954	-84.6121	-59.4350	-32.5765
565	39.7470	-28.1683	65.5823	-71.9425	-51.5034	-27.1971
566	27.3820	-17.4985	44.7956	-46.6123	-35.0795	-16.7425
567	40.8038	-29.8779	67.5119	-75.3041	-53.0714	-28.9902
568	7.0449	-3.0436	11.2004	-9.5928	-8.7040	-2.7229

3.4.6. Feature-Fused Autoencoder Neural Network

A new structure of auto-encoder, namely, Feature Fused Auto Encoder Neural Network, is established based on conventional auto-encoder symmetric properties. Compared with the traditional auto-encoder, the new fused auto-encoder is generated by folding the traditional structure's right side to the left side. This feature fused auto-encoder contains Li1/2 hidden layers similar to a traditional auto-encoder with L hidden layers. However, the variation is the computational cost among the two structures presented in the number of tuned weights, limited to half in the newly developed auto-encoder.

The auto-encoder first compresses the input vector into a lower-dimensional space and then tries to reconstruct the output by minimizing the reconstruction error. The auto-encoder tries to reconstruct the output vector as similarly as possible to the input layer. The dimension reduction for this step is shown in Table (3-4)

Table 3.4. The dimension reduction

	44	45	46	47	48	49
561	14.3001	-0.4516	-0.8065	-17.9619	-1.8264	1.1095
562	19.6854	2.4858	1.8042	-35.1704	1.8334	-0.3682
563	27.5454	9.8153	3.2841	-63.2748	8.9210	-2.9062
564	27.8862	11.6794	3.2567	-68.9199	10.8824	-3.4352
565	26.1930	9.2128	3.1337	-59.8287	8.3589	-2.7300
566	21.0724	4.5849	2.3733	-41.1334	3.8354	-1.3648
567	25.2639	10.2672	2.9744	-61.5573	9.5332	-3.0246
568	7.8046	-0.2491	-0.3452	-9.8743	-0.9279	00.5866

3.4.7. Splitting into training and testing

In this step, the dataset is split into 20% test and 80% train as shown in table 3-5.

Table **3**.5. Train and test data

	0	1	2	3	4
51	-0.7731	-2.4562	-1.2722	0.1163	0.0267
52	0.2542	-4.3567	-1.4030	1.0526	0.5345
53	-0.4241	-1.7968	-0.8442	-0.2696	0.8561
54	-0.4833	-1.9848	-0.9322	-0.2743	0.0599
55	-0.4061	-1.7494	-0.8139	-0.2760	0.0507
56	0.0387	-1.9915	-0.1486	-1.1858	-0.4893

45	46	47	48	49
0.8476	0.2048	-19.8999	-0.3772	-0.6468
3.9874	1.99446	-33.3133	3.4224	-1.2167
4.9385	2.7828	-51.0767	3.8754	-1.2809
1.25116	1.29133	-31.4882	1.0264	-0.1580
9.32249	2.40488	-52.7459	8.8028	-2.7376
0.263552	0.5673	-19.6349	0.1919	0.4082

-

	y_train
505	0
506	1
507	1
508	0
509	0
510	0
511	0

	y_test
50	0
51	1
52	1
53	0
54	0
55	0
56	0

3.4.8. Weight-Tuned Cross-Validation Decision Tree

A decision tree is regarded as generally classifying the divide-and-conquer approach. Based on this model, the data is represented in a tree model, the different characteristics are represented with internal nodes, and leaf nodes are expressed from data sample labels. The tree's path is between the leaf and the root to find the corresponding dataset. The leaf node consists of the final grading outcome, as shown in Figure 3-7.

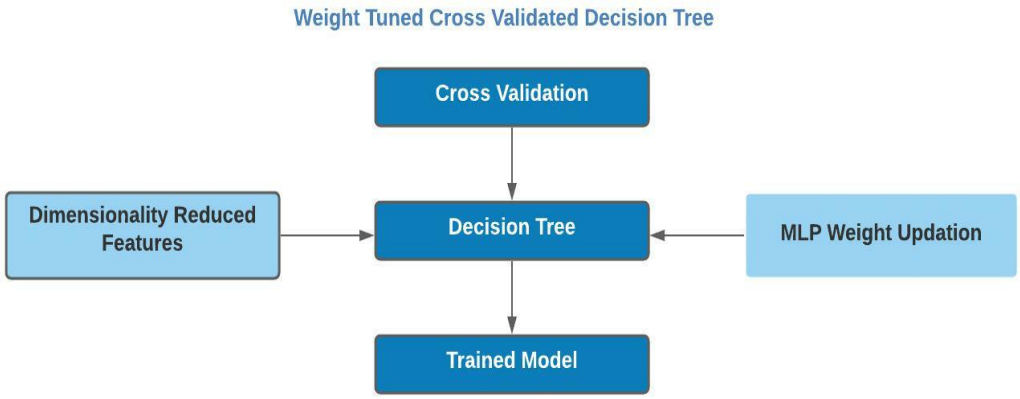


Figure 3.8. Proposed Weight-Tuned Cross-Validation Decision Tree

3.4.8.1. Cross-validation

An essential step when building any machine learning model is to determine how to assess its final performance. To obtain an unbiased measure of the model's performance, we must evaluate it on the data that was not utilized for training. The most straightforward approach to split the data is to employ the train-test split technique. It randomly divides the dataset into two sets (known as training and test sets) so that the specified percentage of the whole dataset is in the training set.

Then, we feed our machine learning model with the training set and assess its performance with the test set. This approach ensures the samples used for training are not used for evaluation and vice versa.

However, the train-split method is not without drawbacks. If the dataset is small, this method is likely to produce inaccurate results. Because of the random partition, the results for different test sets can be drastically different because, in some partitions, samples that are easy to classify are included in the test set, while in others, the 'difficult' ones are included. To address this problem, we employ cross-validation to assess the performance of a machine-learning model. We do not divide the dataset into training and test sets only once in cross-validation. Instead, we divide the dataset into smaller groups and average the performance in each one. This reduces the effect of partition randomness on the results. Numerous cross-validation methods specify various divisions of the available dataset. The k-fold and leave-one-out strategies are the two that are most frequently employed.

In k-fold cross-validation, we initially split up our dataset into k subgroups of equal size. The train-test process is then repeated k times, with each repetition using one of the k subsets as a test set and the remaining k-1 subsets as a training set. By averaging the results across the k trials, we finally calculate an estimate of the model's performance.

In the leave-one-out (LOO) cross-validation, we train our machine-learning model on the amount of our dataset. Every time, we train our model using the remaining samples, and only one sample is used as a test set.

An important factor when choosing between the k-fold and the LOO cross-validation methods is the size of the dataset.

When the size is small, LOO is more appropriate since it will use more training samples in each iteration. That will enable our model to learn better representations.

4. RESULTS AND DISCUSSIONS

4.1. Introduction

This chapter deals with the results and their discussion. The results presented here are the confusion matrix, the ROC curve, the AUC value, and performance parameters. These results have been obtained at different hyperparameters, like weights, train-test size percentages, batch size, Hidden units, and dropout. The results were finally compared with other studies.

4.2. Results overview at different hyperparameters

4.2.1. Weights effects

The results in Table 4-1 and Figure 4-1 illustrate the effect of varying weights on model accuracy using the WDBC dataset across different test sizes. Ten different weights were tested, and the results showed a generally positive, fluctuating relationship between increasing weights and improving classification performance.

At low weight levels, the models performed moderately well. For example, at a weight of 1 and a 10% test size, the accuracy reached 0.8596, one of the lowest values recorded. As the weights increased, the accuracy values showed significant improvements, albeit with some fluctuations depending on the test size. For example, at a weight of 5 and a 15% test size, the accuracy reached 0.9534, while at the same weight and a 30% test size, it dropped to 0.9298. This indicates that increasing weights often improves performance, but it does not follow a consistent linear pattern, as it remains sensitive to the effect of test size.

The most notable result was recorded at weight (9) with a 25% test size, where the model achieved the highest recorded accuracy of 0.9580. This demonstrates that the interaction between weight tuning and test size can provide optimal conditions for the model to generalize efficiently, by balancing underfitting at low weights and overfitting at very high weights.

It was also found that at large test sizes (30%–45%), fluctuations in accuracy became more pronounced. While some weights maintained relatively stable performance (e.g., weight 10 with a 45% test size, which achieved 0.9338), others experienced a significant decline (e.g., weight 9 with a 30% test size, which dropped to 0.9123). This discrepancy reflects that the model's robustness depends in part on optimizing the weights to match the data split.

Overall, these results highlight that careful tuning of weight parameters is essential for improving model accuracy. Intermediate weights (such as 7 and 9) achieved superior results across multiple test sizes, underscoring their role in achieving stability. These results are consistent with previous literature that has emphasized the importance of improving regularization parameters in enhancing the efficiency of breast cancer classification models, whether in statistical models or machine learning-based algorithms.

Table **4**.1. Accuracy at different weights and different test sizes

Weight	accuracy 10%test	accuracy 15%test	accuracy 20%test	accuracy 25%test	accuracy 30%test	accuracy 35%test	accuracy 40%test	accuracy 45%test
1	0.8771	0.9418	0.9210	0.9440	0.9532	0.9300	0.9122	0.9338
2	0.8596	0.9418	0.9298	0.9370	0.9473	0.9200	0.9035	0.9182
3	0.8771	0.9418	0.9298	0.9370	0.9415	0.9300	0.9078	0.9182
4	0.8596	0.9418	0.9298	0.9300	0.9532	0.9200	0.9166	0.9221
5	0.8596	0.9534	0.9210	0.9300	0.9298	0.9300	0.8947	0.9143
6	0.8596	0.9418	0.9298	0.9300	0.9298	0.9050	0.9078	0.9143
7	0.8771	0.9534	0.9298	0.95104	0.9415	0.9250	0.9122	0.9143
8	0.8771	0.9302	0.9210	0.93706	0.9415	0.9350	0.9254	0.9105
9	0.8771	0.9418	0.9210	0.9580	0.9122	0.9200	0.9078	0.9182
10	0.8596	0.9302	0.9298	0.9370	0.9356	0.9300	0.9122	0.9338

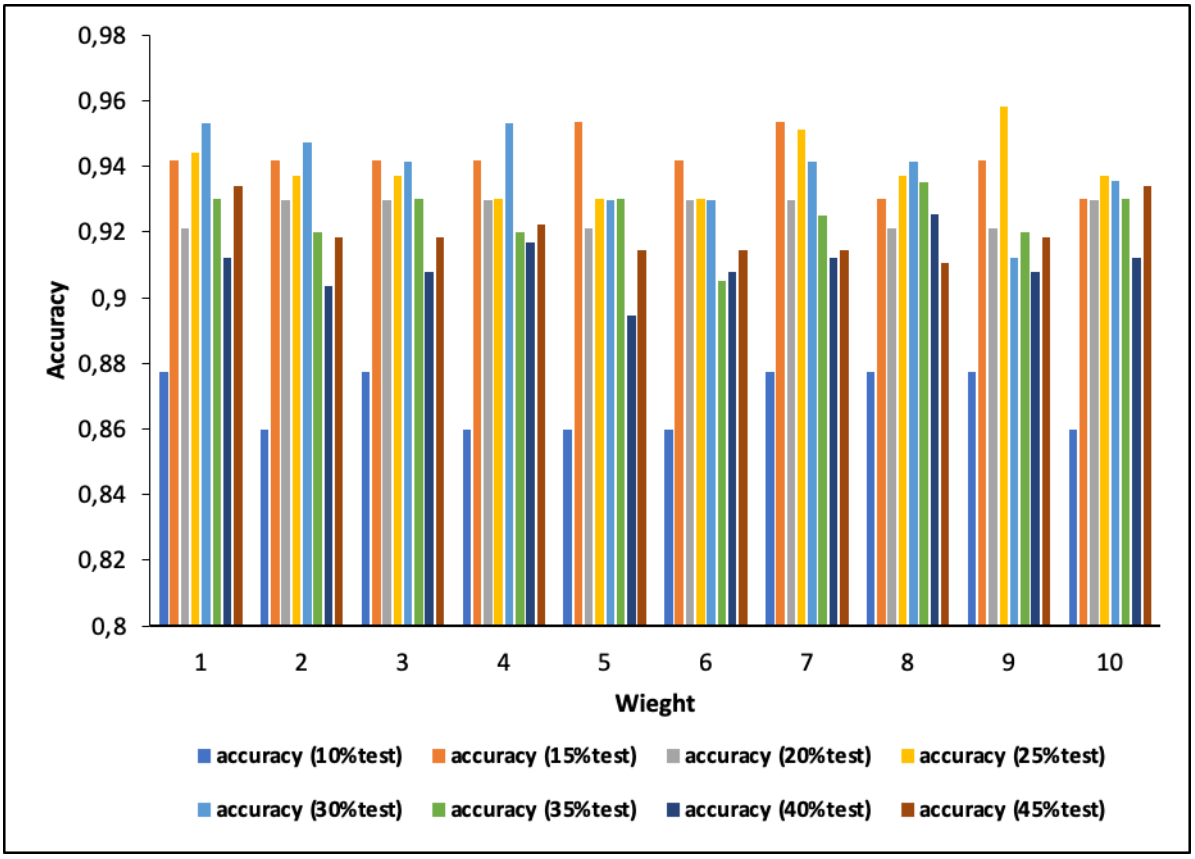


Figure 4.1. Accuracy vs weight

4.2.2. Train and Test sizes effects

These performance parameters are evaluated at different iterations of a dataset. The iterations that were used were (10:90, 15:85, 20:80, 25:75, 30:70, 35:65, 40:60, and 45:55) for testing and training data, respectively. Table (4-2) shows the accuracy for each iteration. Figure (4-2) explains the accuracy at different weights for each iteration. Even with a decrease in the proportion of training data, the network can deliver an accuracy of classification higher than 90%.

Table **4**.2. The accuracy at different iterations and different train and test sizes.

Iteration	Test size %	Train size %	Accuracy
1	10	90	0.9947
2	15	85	0.9929
3	20	80	0.9824
4	25	75	0.9824
5	30	70	0.9806
6	35	65	0.9683
7	40	60	0.9648
8	45	55	0.9666

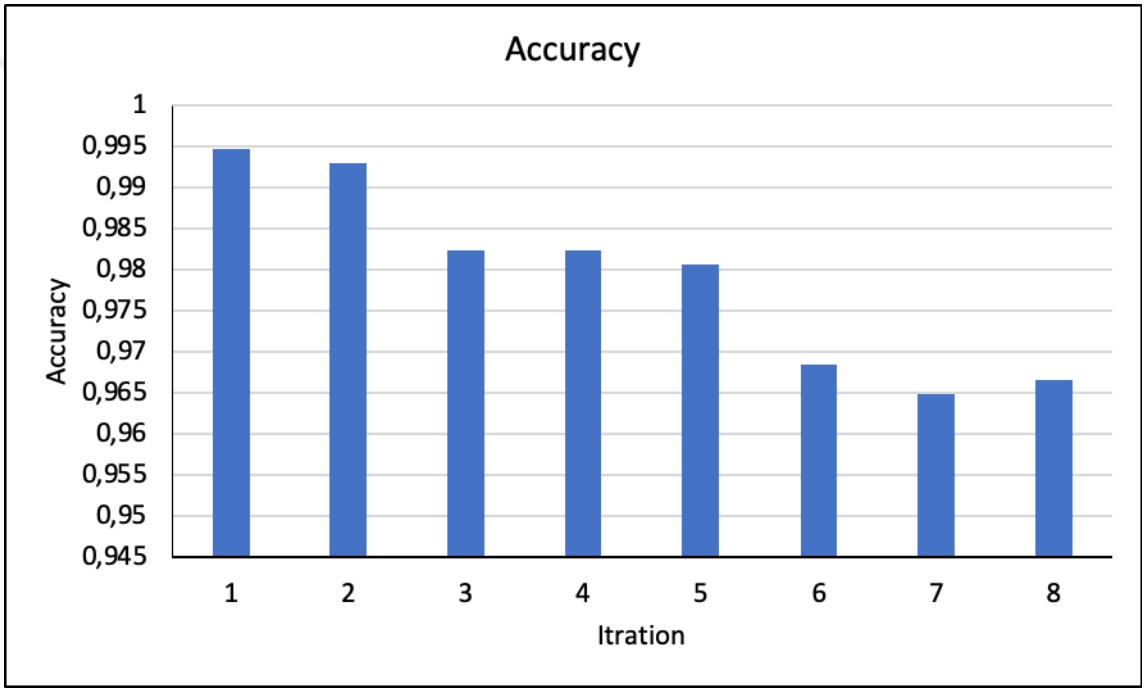


Figure 4.2. Accuracy at different weights for different test sizes

The projected result at 0.10 test size is displayed in the confusion matrix, as well as the model's computed performance. The total number of correct predictions is 558, with 11 incorrect forecasts. Figure (4-3) shows the ROC curve of the Weight-Tuned Cross-Validated Decision Tree. The accuracy under the curve is 99.6% for the Weight-Tuned Cross-Validated Decision Tree classifier. The confusion matrix indicates that the model correctly classified 350 attack samples, while incorrectly classifying 7 attacks as normal traffic. For normal traffic, 208 samples were correctly classified, while 4 were incorrectly identified as attacks. These results demonstrate that the model performs reliably in both categories, despite a small number of false negatives in the attack category and a limited number of false positives in the normal category.

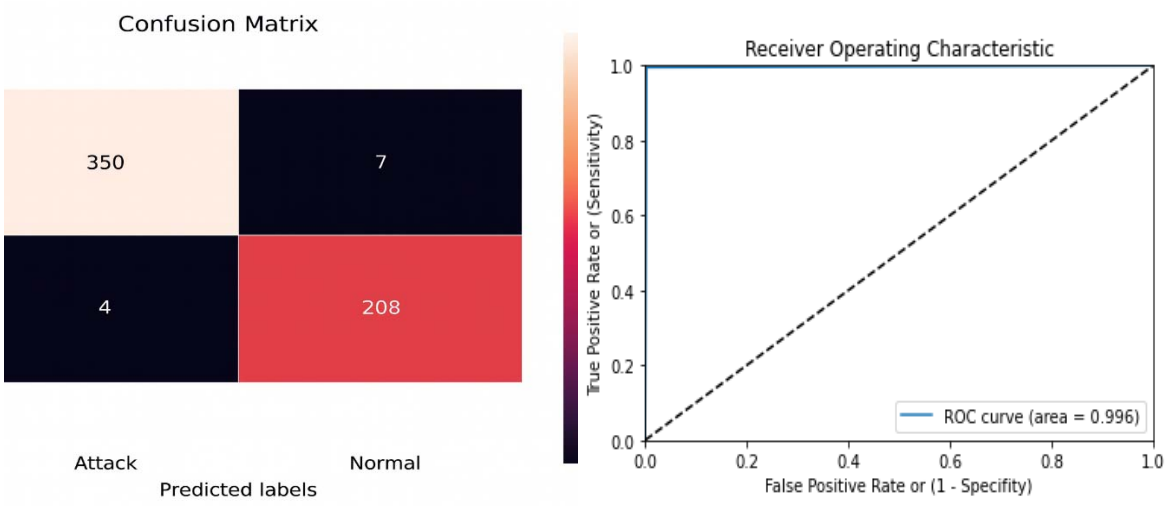


Figure 4.3. The confusion matrix and the ROC curve for the iteration with a test size of 10% and train size of 90%

The projected result at 0.15 test size is displayed in the confusion matrix, as well as the model's computed performance. The total number of correct predictions is 563, with 6 incorrect forecasts. Figure (4-4) shows the ROC curve of the Weight-Tuned Cross-Validated Decision Tree. The accuracy under the curve is 98.2% for the Weight-Tuned Cross-Validated Decision Tree classifier. The confusion matrix in the final model shows that 357 samples were correctly classified as attack traffic, while none were classified as normal traffic. For normal traffic, the model correctly classified 206 samples, while misclassifying 6 as attacks. This indicates that the model demonstrates strong attack detection performance with no false negatives for the attack category, while maintaining a low false positive rate for normal traffic.

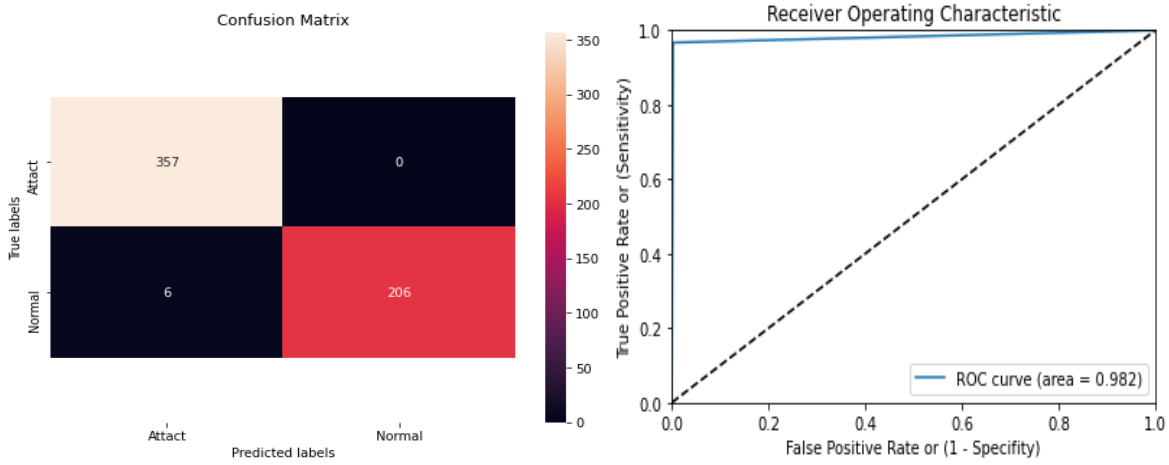


Figure 4.4. The confusion matrix and the ROC curve for the iteration with a test size of 15% and a train size of 85%

The projected result at 0.20 test size is displayed in the confusion matrix, as well as the model's computed performance. The total number of correct predictions is 559, with 10 incorrect forecasts. Figure (4-5) shows the ROC curve of the Weight-Tuned Cross-Validated Decision Tree. It shows that the accuracy under the curve is 98.7% for the Weight-Tuned Cross-Validated Decision Tree classifier. According to the confusion matrix, the model correctly classified 355 attack samples and incorrectly classified only 2 as normal traffic. For normal traffic, 204 samples were correctly classified, while 8 were incorrectly predicted as attacks. These results indicate that the model maintains high detection power with very low false-negative rates for attack samples and few false-positive rates for normal traffic. Overall, the model demonstrates stable and robust performance in both categories.

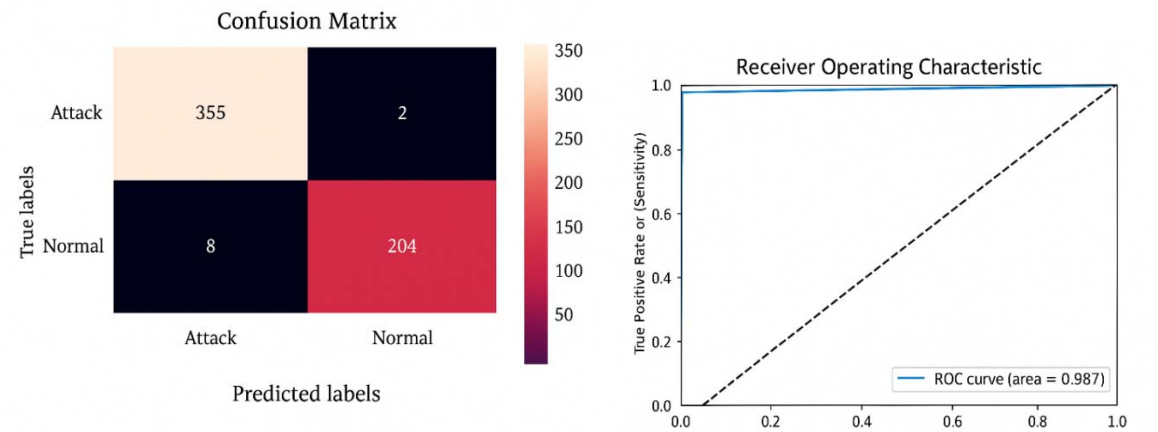


Figure 4.5. The confusion matrix and the ROC curve for the iteration with a test size of 20% and a train size of 80%

The projected result at 0.25 test size is displayed in the confusion matrix, as well as the model's computed performance. The total number of correct predictions is 559, with 10 incorrect forecasts. Figure (4-6) shows the ROC curve of the decision tree. It shows that the accuracy under the curve is 97.8% for the decision tree classifier. The confusion matrix shows that the model correctly classified 354 attack samples, while incorrectly classifying 3 samples as normal. For the normal category, the model correctly identified 205 samples and incorrectly classified 7 as attacks. These results indicate that the model performs well with a low number of false negatives for attack samples and a relatively small number of false positives for normal traffic. This reflects consistent and reliable classification behavior in both categories.

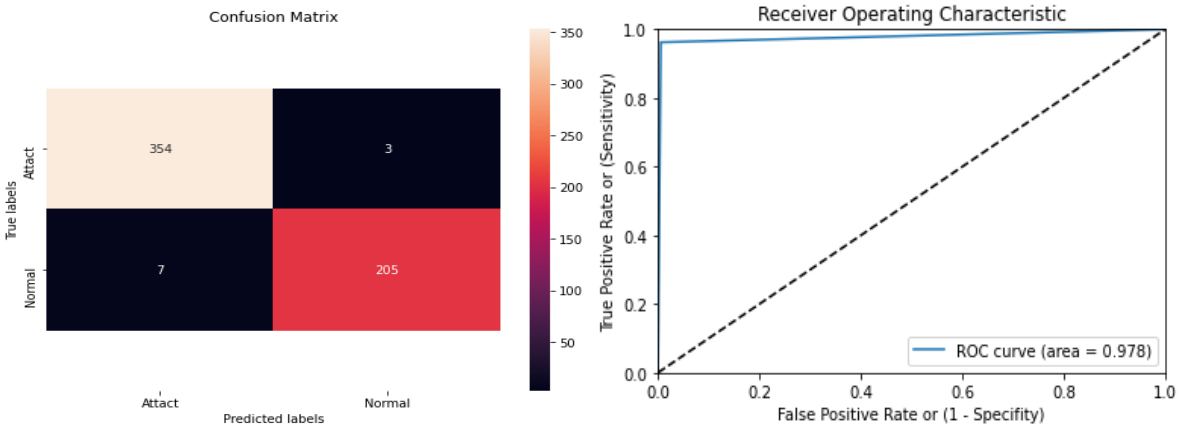


Figure 4.6.The confusion matrix and the ROC curve for the iteration with a test size of 25% and a train size of 75%

The projected result at 0.30 test size is displayed in the confusion matrix, as well as the model's computed performance. The total number of correct predictions is 558, with 11 incorrect forecasts. Figure (4-7) shows the ROC curve of the Weight-Tuned Cross-Validated Decision Tree. The accuracy under the curve is 97.8% for the Weight-Tuned Cross-Validated Decision Tree classifier. The confusion matrix indicates that the model correctly classified 352 attack samples and incorrectly classified 5 samples as normal. For the normal category, the model correctly identified 206 samples and incorrectly classified 6 as attacks. These results reflect stable model performance with low false negatives for attack traffic and low false positives for normal data traffic.

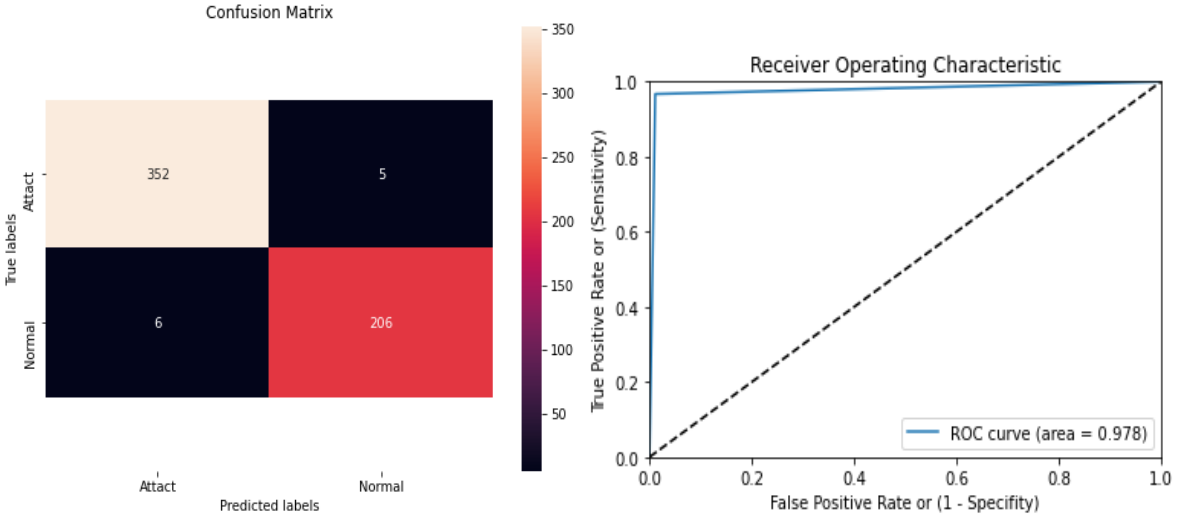


Figure 4.7. The confusion matrix and the ROC curve for the iteration with a test size of 30% and a train size of 70%

The projected result at 0.35 test size is displayed in the confusion matrix, as well as the model's computed performance. The total number of correct predictions is 551, with 18 incorrect forecasts. Figure (4-8) shows the ROC curve of the Weight-Tuned Cross-Validated Decision Tree. The accuracy under the curve is 97.7% for the Weight-Tuned Cross-Validated Decision Tree

classifier. The confusion matrix shows that the model correctly identified 349 attack samples, while 8 attacks were incorrectly classified as normal traffic. Regarding normal traffic, 202 samples were correctly classified, while 10 were incorrectly predicted as attacks. These results indicate a slight increase in misclassification rates compared to previous runs, but still reflect generally reliable performance in both categories.

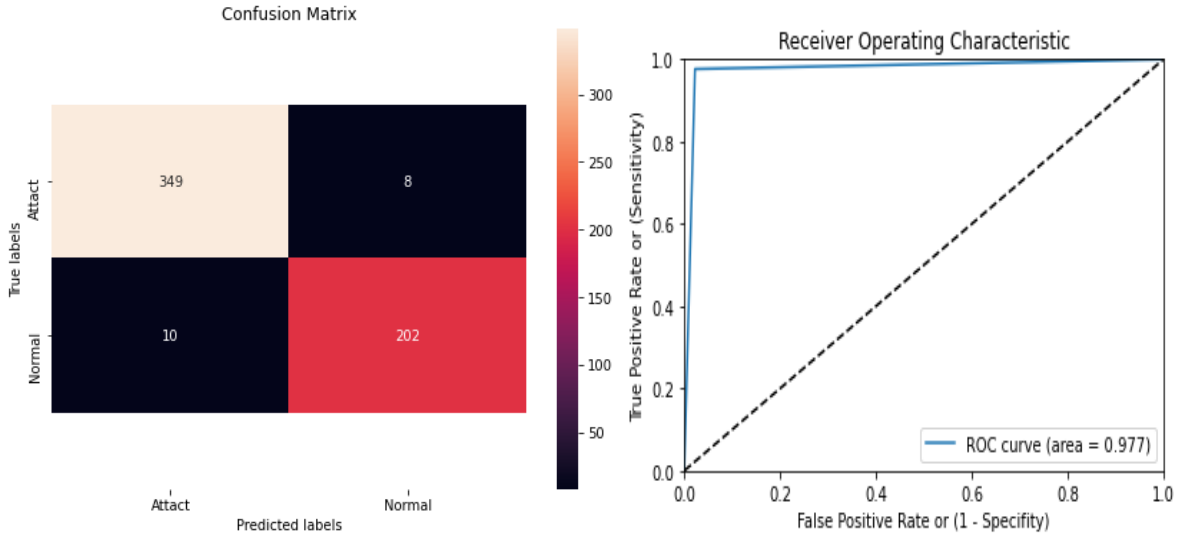


Figure 4.8. The confusion matrix and the ROC curve for the iteration with a test size of 35% and a train size of 65%

The projected result at 0.40 test size is displayed in the confusion matrix, as well as the model's computed performance. The total number of correct predictions is 449, with 20 incorrect forecasts. Figure (4-9) shows the ROC curve of the Weight-Tuned Cross-Validated Decision Tree. The accuracy under the curve is 96.8% for the Weight-Tuned Cross-Validated Decision Tree classifier. The confusion matrix shows that the model correctly classified 351 attack samples, while 6 samples were incorrectly classified as normal. For normal traffic, 198 samples were correctly identified, while 14 were incorrectly predicted as attacks. This configuration shows slightly higher false positives for the normal category compared to previous runs; however, the model still demonstrates strong and consistent performance in detecting attack traffic with relatively low false negative rates.

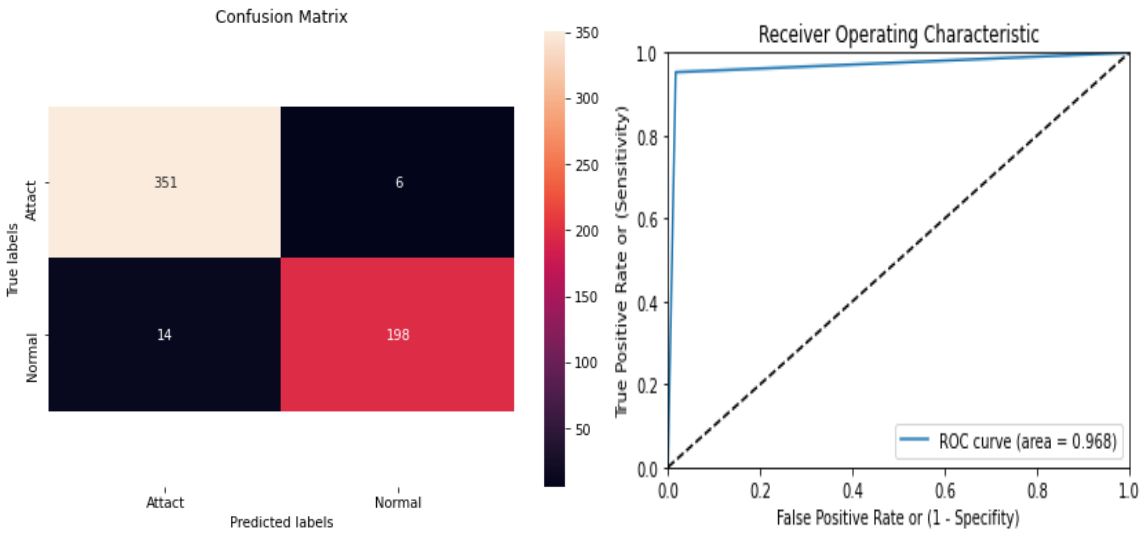


Figure 4.9. The confusion matrix and the ROC curve for the iteration with a test size of 40% and a train size of 60%

The projected result at 0.45 test size is displayed in the confusion matrix, as well as the model's computed performance. The total number of correct predictions is 550, with 19 incorrect forecasts. Figure (4-10) shows the ROC curve of the Weight-Tuned Cross-Validated Decision Tree. The accuracy under the curve is 96.5% for the Weight-Tuned Cross-Validated Decision Tree classifier. The confusion matrix shows that the model correctly classified 347 attack samples, while 10 attack samples were incorrectly classified as normal. For the normal category, 203 samples were correctly predicted, and 9 were incorrectly classified as attacks. Compared to previous results, this run shows a slightly higher false negative rate for attack traffic, although it maintains stable classification performance for the normal category. Overall, the model continues to demonstrate reliable behavior in both categories.

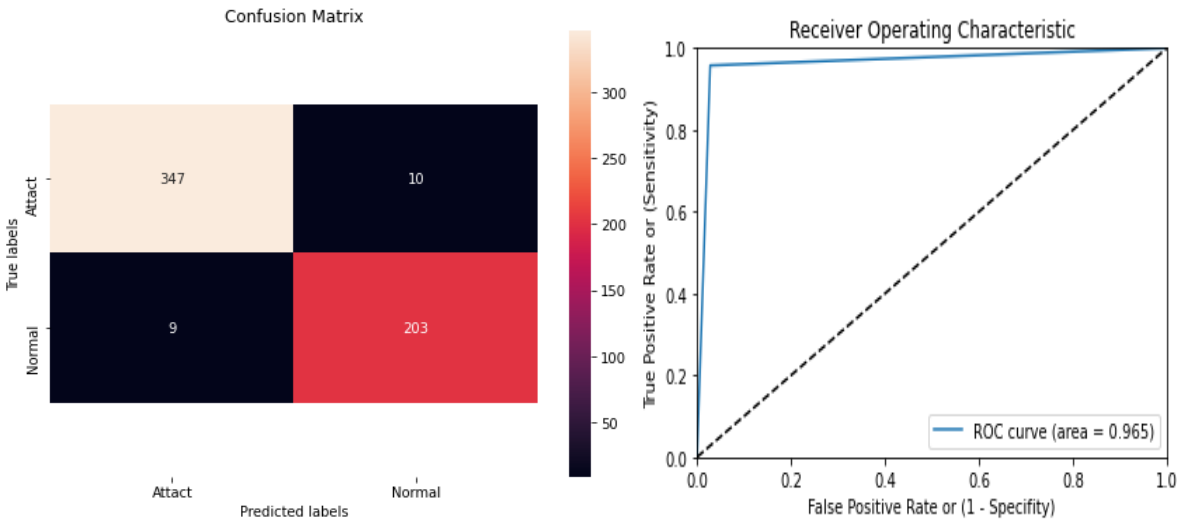


Figure 4.10. The confusion matrix and the ROC curve for the iteration with a test size of 45% and a train size of 55%

Table **4**.3. Classification report at 10% test size

		precision	recall	f1-score	support
	0	0.9900	0.9800	0.9800	357
	1	0.9700	0.9800	0.9700	212
Accuracy				0.9800	569
Macro avg		0.9800	0.9800	0.9800	569
Weighted avg		0.9800	0.9800	0.9800	569

In the table (4-3), the overall F1-score achieved in this case at 10%test size is 98%. The individual F1-score is 98% for benign and 97% for malignant.

Table 4.4. Classification report at 15% test size

		precision	recall	F1-score	support
	0	0.9900	1.0000	0.9900	357
	1	1.0000	0.9800	0.9900	212
Accuracy				0.9900	569
Macro avg		0.9900	0.9900	0.9900	569
Weighted avg		0.9900	0.9900	0.9900	569

In the table (4-4), the overall F1-score achieved in this case at a 15% test size is 99%. The individual F1-score is 99% for benign and 99% for malignant.

Table **4**.5. Classification report at 20% test size

		precision	recall	F1-score	support
	0	0.9800	0.9900	0.9900	357
	1	0.9900	0.9600	0.9800	212
Accuracy				0.9800	569
Macro avg		0.9800	0.9800	0.9800	569
Weighted avg		0.9800	0.9800	0.9800	569

In the table (4-5), the overall F1-score achieved in this case at a 20% test size is 98%. The individual F1-score is 99% for benign and 98% for malignant.

Table **4**.6. Classification report at 25% test size

		precision	recall	f1-score	support
	0	0.9800	0.9900	0.9900	357
	1	0.9900	0.9700	0.9800	212
Accuracy				0.9800	569
Macro avg		0.9800	0.9800	0.9800	569
Weighted avg		0.9800	0.9800	0.9800	569

In the table (4-6), the overall F1-score achieved in this case at a 25% test size is 98%. The individual F1-score is 99% for benign and 98% for malignant.

Table **4**.7.Classification report at 30% test size

		precision	recall	f1-score	support
	0	0.9800	0.9900	0.9800	357
	1	0.9800	0.9700	0.9700	212
Accuracy				0.9800	569
Macro avg		0.9800	0.9800	0.9800	569
Weighted avg		0.9800	0.9800	0.9800	569

In the table (4-7), the overall F1-score achieved in this case at a 30% test size is 98%. The individual F1-score is 98% for benign and 97% for malignant.

Table **4**.8. Classification report at 35% test size

		precision	recall	f1-score	support
	0	0.9700	0.9800	0.9700	357
	1	0.9600	0.9500	0.9600	212
Accuracy				0.9700	569
Macro avg		0.9700	0.9700	0.9700	569
Weighted avg		0.9700	0.9700	0.9700	569

In the table (4-8), the overall F1-score achieved in this case at a 35% test size is 97%. The individual F1-score is 97% for benign and 96% for malignant.

Table **4**.9. Classification report at 40% test size

		precision	recall	F1-score	support
	0	0.9600	0.9800	0.9700	357
	1	0.9700	0.9300	0.9500	212
Accuracy				0.9600	569
Macro avg		0.9700	0.9600	0.9600	569
Weighted avg		0.9600	0.9600	0.9600	569

In the table (4-9), the overall F1-score achieved in this case at a 40% test size is 96%. The individual F1-score is 97% for benign and 95% for malignant.

Table **4**.10. Classification report at 45% test size

		precision	recall	F1-score	support
	0	0.9700	0.9700	0.9700	357
	1	0.9500	0.9600	0.9600	212
Accuracy				0.9700	569
Macro avg		0.9600	0.9600	0.9600	569
Weighted avg		0.9700	0.9700	0.9700	569

In the table (4-10), the overall F1-score achieved in this case at a 45% test size is 97%. The individual F1-score is 97% for benign and 96% for malignant.

Table **4**.11. AUC at different iterations and different test sizes

Iteration	**Test size %**	**Train size %**	**AUC**
1	10	90	0.9960
2	15	85	0.9820
3	20	80	0.9870
4	25	75	0.9780
5	30	70	0.9780
6	35	65	0.9770
7	40	60	0.9680
8	45	55	0.9650

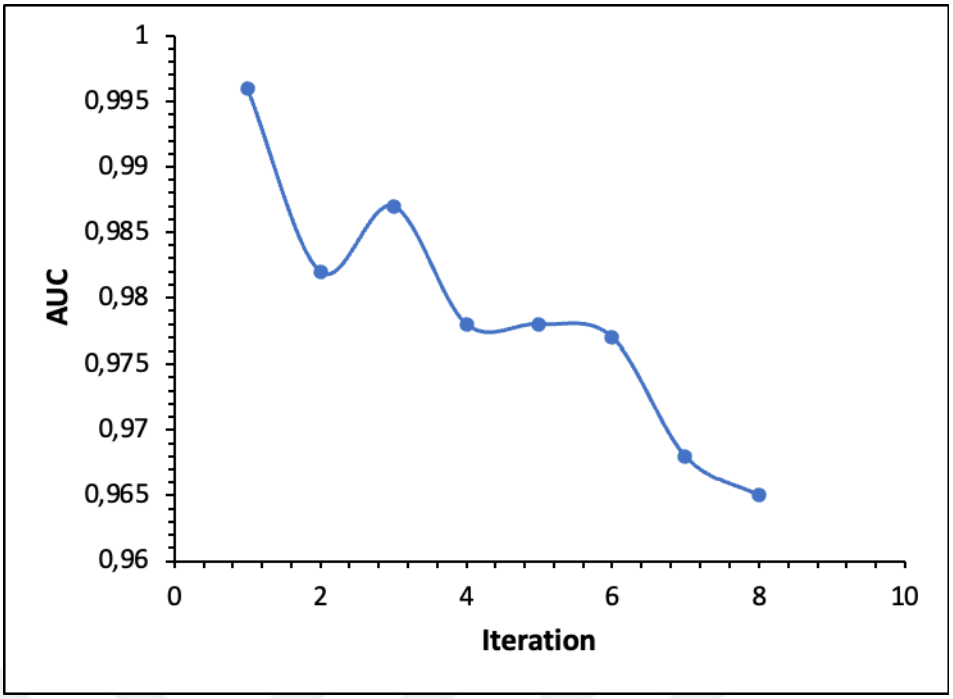


Figure 4.11. AUC vs Iteration

As shown in the figure (4-11) and table (4-11), the AUC decreased gradually as the test size increased, meaning that using less training data reduced the model's learning capacity and generalization performance. The highest AUC (0.996) was achieved with 90% training and 10% testing, indicating optimal learning at that ratio. The stability between iterations 4–6 suggests that the model maintains reasonable robustness until the training data falls below 65

Choosing an appropriate train–test ratio is crucial for reliable model evaluation. For this diagnostic model, a 90–10 or 80–20 split provides a good balance between training sufficiency and testing reliability, achieving near-perfect AUC.

4.2.3. Batch-size effects

At a hidden unit equal to 256 and dropout equal to 0.45, different values of batch size (32, 64, 128, and 256) were used to get the optimal batch size at which the performance parameters improve. Tables from (4-12) to (4-16) explain the classification reports and the ROC curves with the AUC values. Figure from (4-12) to (4-15) displays the performance parameters (accuracy, precision, recall, and F1-score) as well as AUC at different values of batch size (32, 64, 128, and 256). The performance parameters and AUC have higher values at the batch size of 128. Then 128 represents the optimal value of the batch size in this work.

Table 4.12. Classification report at Batch size 32

		precision	recall	F1-score	support
	0	0.9900	1.0000	0.9900	357
	1	1.0000	0.9900	0.9900	212
Accuracy				0.9900	569
Macro avg		0.9900	0.9900	0.9900	569
Weighted avg		0.9900	0.9900	0.9900	569
Accuracy score		0.9912			

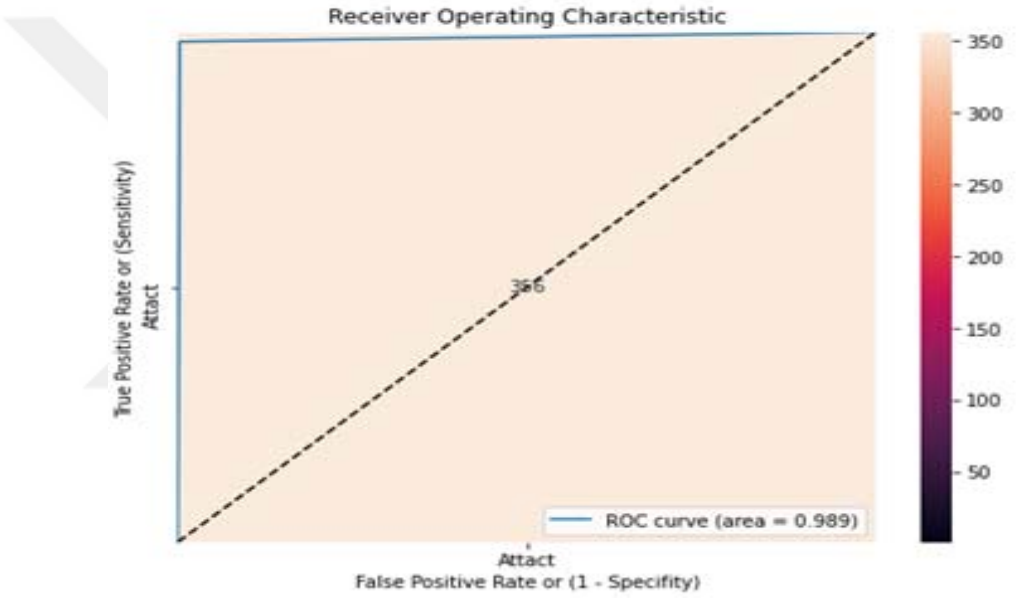


Figure 4.12.The ROC curve for the Batch size 32

Table **4**.13. Classification report at Batch size 64

		precision	recall	F1-score	Support
	0	0.9900	0.9900	0.9900	357
	1	0.9900	0.9900	0.9900	212
Accuracy				0.99	569
Macro avg		0.9900	0.9900	0.9900	569
Weighted avg		0.9900	0.9900	0.9900	569
Accuracy score		0.9929			

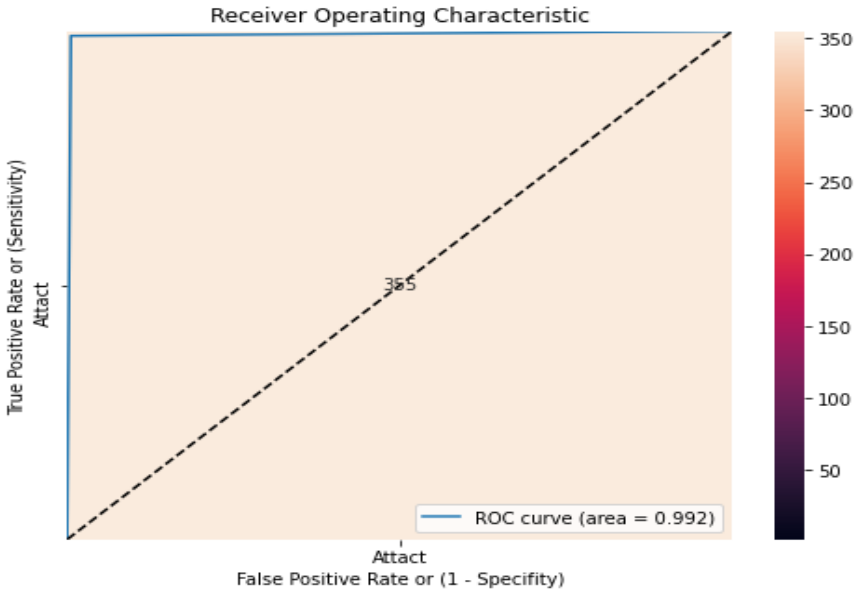


Figure 4.13. The ROC curve for the Batch size 64

Table **4**.14. Classification report at Batch size 128

		precision	recall	F1-score	Support
	0	0.9900	1.0000	1.0000	357
	1	1.0000	0.9900	0.9900	212
Accuracy				0.9900	569
Macro avg		0.9900	0.9900	0.9900	569
Weighted avg		0.9900	0.9900	0.9900	569
Accuracy score		0.9947			

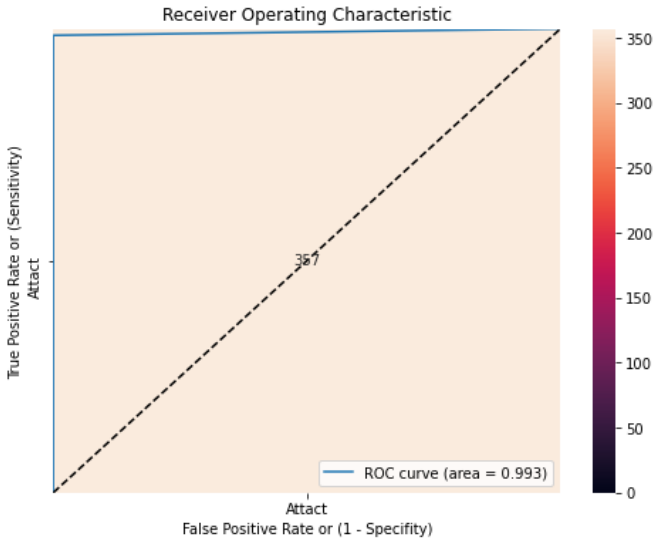


Figure 4.14. The ROC curve for the Batch size 128

Table **4**.15. Classification report at Batch size 256

		precision	recall	F1-score	support
	0	0.9900	0.9900	0.9900	357
	1	0.9900	0.9800	0.9800	212
Accuracy				0.9900	569
Macro avg		0.9900	0.9900	0.9900	569
Weighted avg		0.9900	0.9900	0.9900	569
Accuracy score		0.9896			

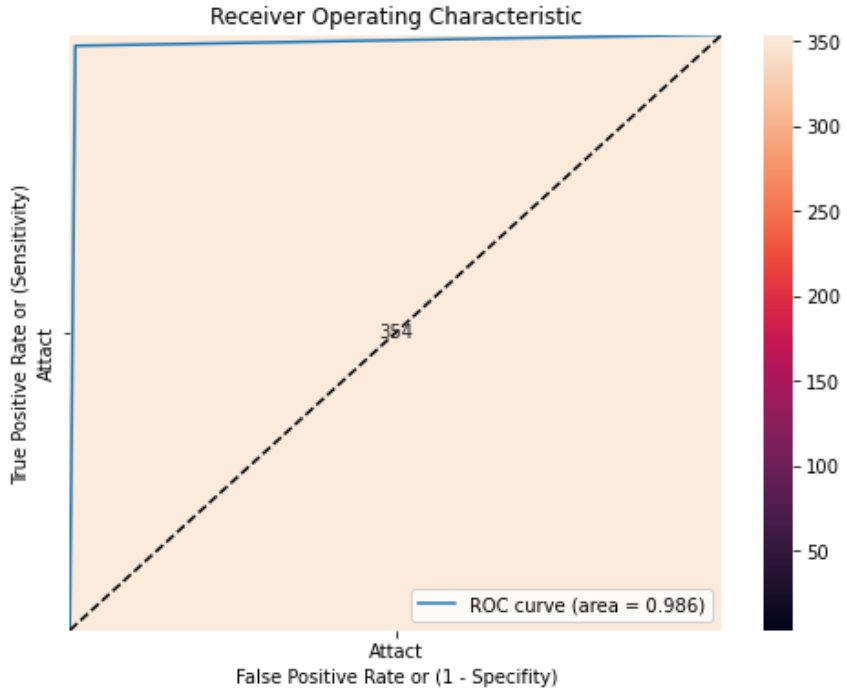


Figure 4.15. The ROC curve for the Batch size 256

Table **4**.1**6**. Performance parameters at different batch sizes

batch size	Accuracy	Class	Precision	recall	F1-score	AUC
32	0.9912	B	0.9900	1.0000	0.9900	0.98
		M	1.0000	0.9800	0.9900	
64	0.9929	B	0.9900	0.9900	0.9900	0.99
		M	0.9900	0.9900	0.9900	
128	0.9947	B	0.9900	1.0000	1.0000	0.99
		M	1.0000	0.9900	0.9900	
256	0.9876	B	0.9900	0.9900	0.9900	0.98
		M	0.9900	0.9800	0.9800	

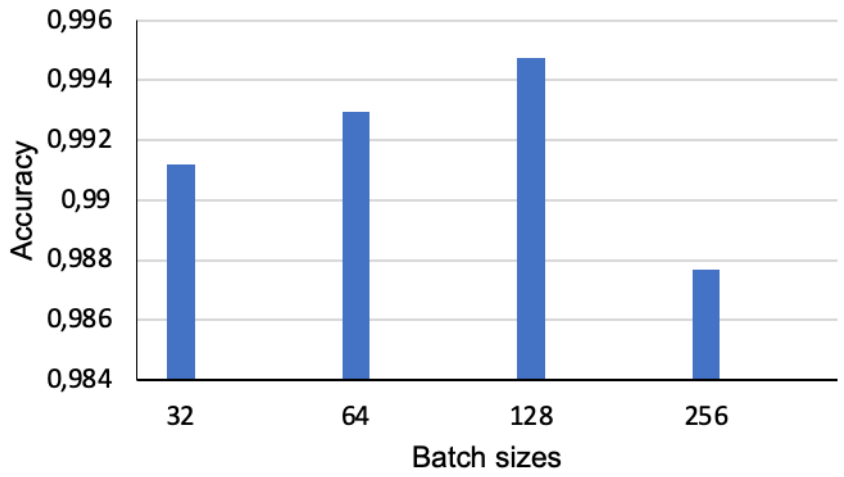


Figure 4.16. Accuracy at different batch sizes

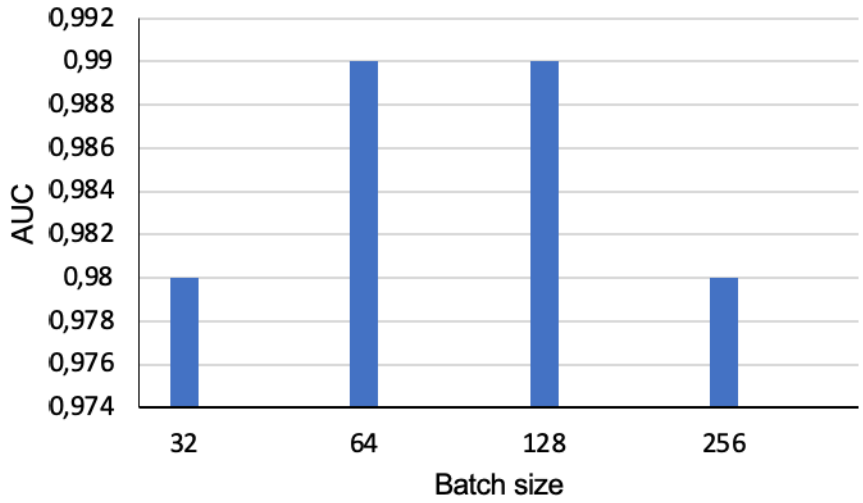


Figure 4.17. The AUC at different batch sizes

As shown in figures (4-16) and (4-17), and as indicated in table (4-16), the accuracy improves as the batch size increases until a certain point, after which it begins to decrease. The improvement results from larger batch sizes, which enable the model to see more diverse examples during each training iteration. This greater exposure to varied samples can help the model learn more robust, generalizable patterns in the data, leading to better performance on unseen data. However, the increase in batch size must be limited at some point, as accuracy can decline beyond that point, since extremely large batch sizes might cause overshooting of the optima and lead the model to converge to a suboptimal solution.

4.2.4. Hidden Unit Effects

Figures from (4-18) to (4-25) show the classification reports and ROC curves at different hidden units (16, 32, 64, 128, and 256)

Table **4**.**17**. Classification report at Hidden Unit 16

		precision	recall	F1-score	support
	0	0.9900	0.9900	0.9900	357
	1	0.9800	0.9800	0.9800	212
Accuracy				0.9900	569
Macro avg		0.9800	0.9800	0.9800	569
Weighted avg		0.9900	0.9900	0.9900	569
Accuracy score		0.9859			

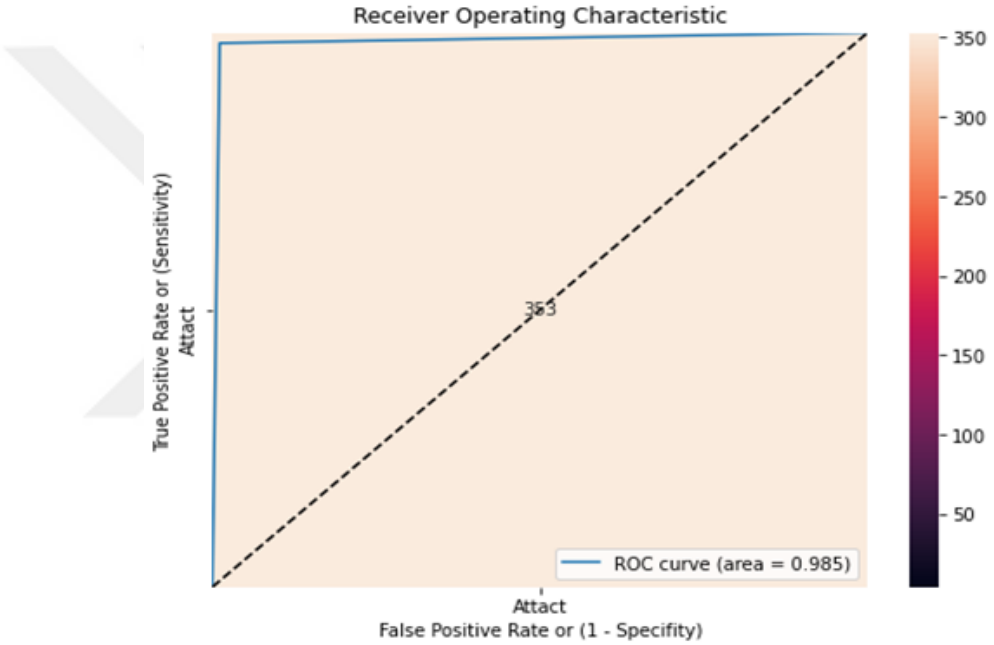


Figure 4.18. The ROC curve for the Hidden Unit 16

Table 4.18. Classification report at Hidden Unit 32

		Precision	recall	F1-score	Support
	0	0.9900	0.9900	0.9900	357
	1	0.9800	0.9800	0.9800	212
Accuracy				0.9900	569
Macro avg		0.9800	0.9800	0.9800	569
Weighted avg		0.9900	0.9900	0.9900	569
Accuracy score		0.9859			

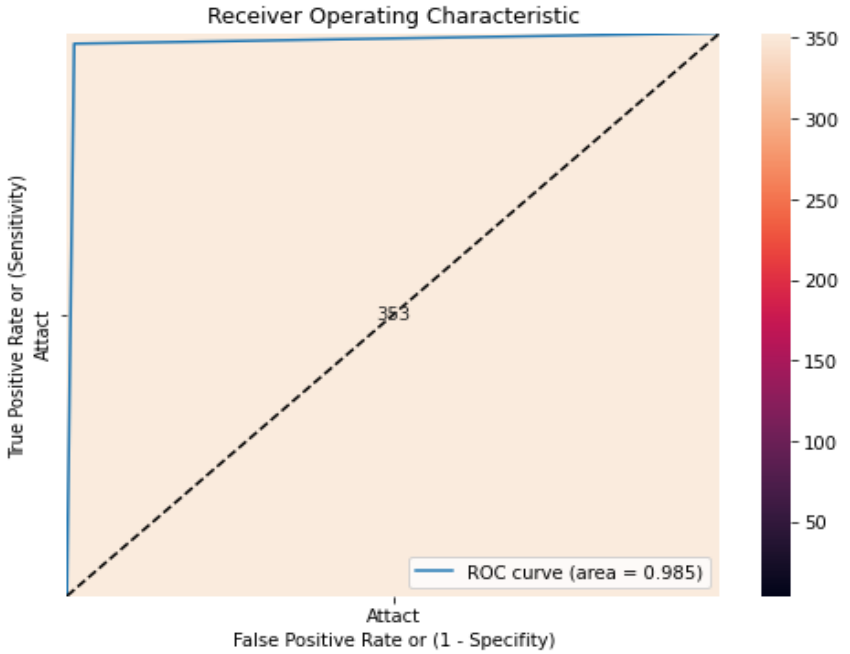


Figure 4.19. The ROC curve for the Hidden Unit 32

Table **4**.**19**. Classification report at Hidden Unit 64

		precision	recall	F1-score	Support
	0	0.9900	0.9900	0.9900	357
	1	0.9900	0.9800	0.9900	212
Accuracy				0.9900	569
Macro avg		0.9900	0.9900	0.9900	569
Weighted avg		0.9900	0.9900	0.9900	569
Accuracy score		0.9455			

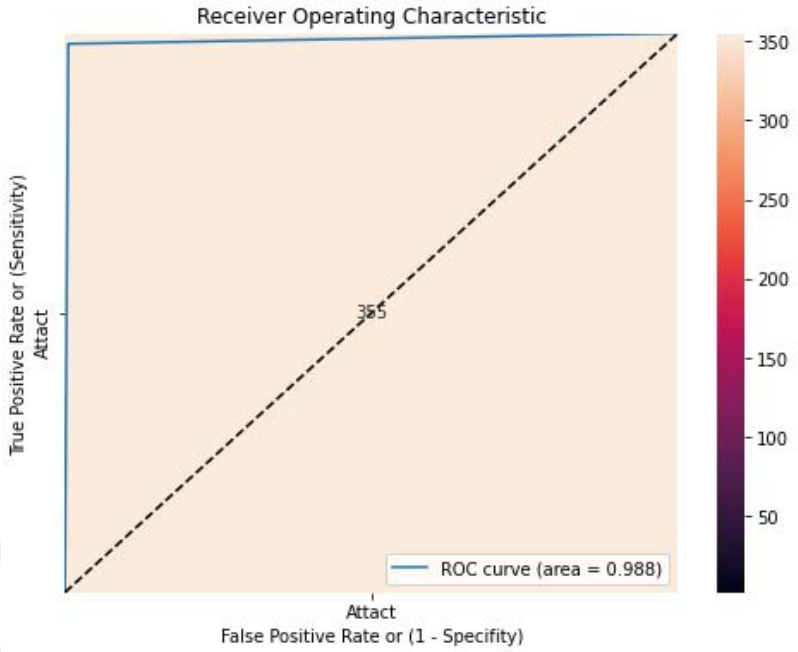


Figure 4.20.The ROC curve for the Hidden Unit 64

Table 4.20. Classification report at Hidden Unit 128

		Precision	recall	F1-score	support
	0	0.9900	0.9900	0.9900	357
	1	0.9900	0.9900	0.9900	212
Accuracy				0.9900	569
Macro avg		0.9900	0.9900	0.9900	569
Weighted avg		0.9900	0.9900	0.9900	569
Accuracy score		0.9894			

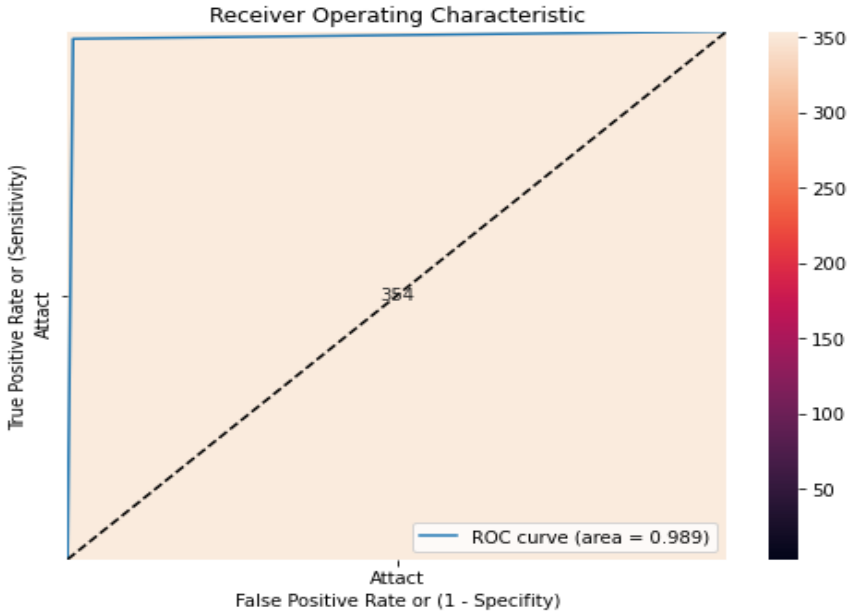


Figure 4.21. The ROC curve for the Hidden Unit 128

Table **4**.2**1**. Classification report at Hidden Unit 256

		Precision	recall	F1-score	support
	0	0.9900	1.0000	0.9900	357
	1	1.0000	0.9800	0.9900	212
Accuracy				0.9900	569
Macro avg		0.9900	0.9900	0.9900	569
Weighted avg		0.9900	0.9900	0.9900	569

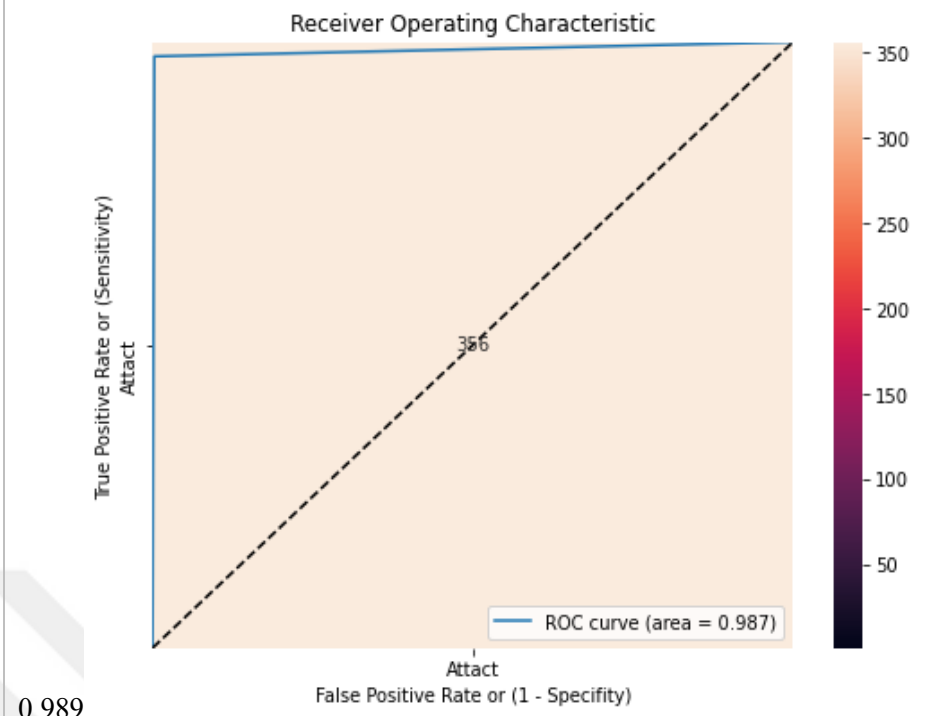
Accuracy score	 0.989 4

Figure 4.22.The ROC curve for the Hidden Unit 256

Table 4.22. Classification report at Hidden Unit 512

		Precision	Recall	F1-score	support
	0	0.9900	1.0000	0.9900	357
	1	1.0000	0.9800	0.9900	212
Accuracy				0.9900	569
Macro avg		0.9900	0.9900	0.9900	569
Weighted avg		0.9900	0.9900	0.9900	569
Accuracy score		0.9912			

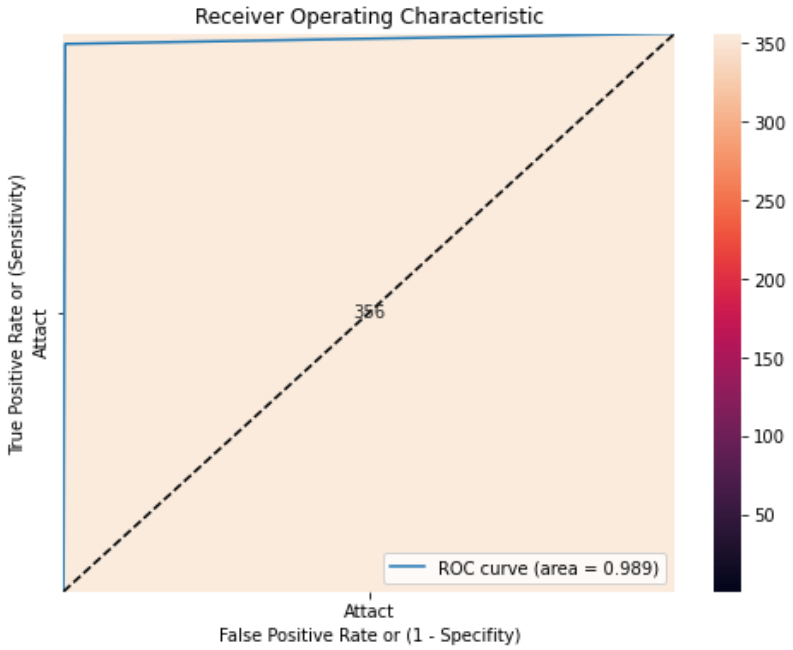


Figure 4.23. The ROC curve for the Hidden Unit 512

Table **4.23**.Performance parameters at different hidden units

Hidden Unit	Accuracy	Class	Precision	recall	F1-score	AUC
16	0.9859	B	0.9900	0.9900	0.9900	0.9850
		M	0.9800	0.9800	0.9800	
32	0.9859	B	0.9900	1.0000	0.9900	0.9850
		M	1.0000	0.9800	0.9900	
64	0.9894	B	0.9900	0.9900	0.9900	0.9880
		M	0.9900	0.9900	0.9900	
128	0.9894	B	0.9900	1.0000	1.0000	0.9890
		M	1.0000	0.9900	0.9900	
256	0.9912	B	0.9900	1.0000	1.0000	0.9900
		M	1.0000	0.9800	0.9900	
512	0.9912	B	0.9900	1.0000	1.0000	0.9890
		M	1.0000	0.9800	0.9900	

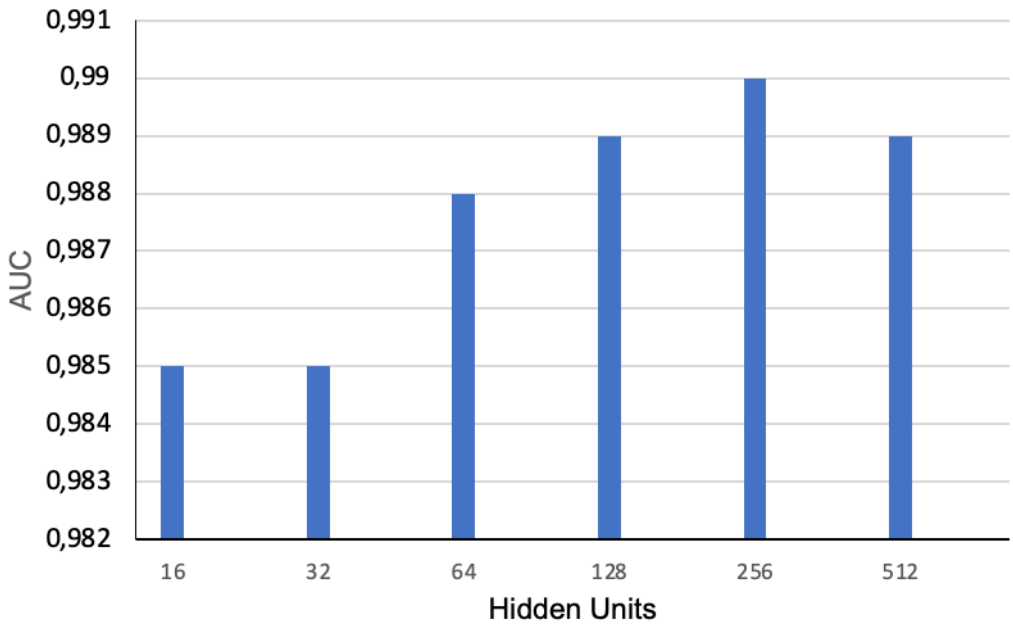


Figure 4.24. AUC at different hidden units

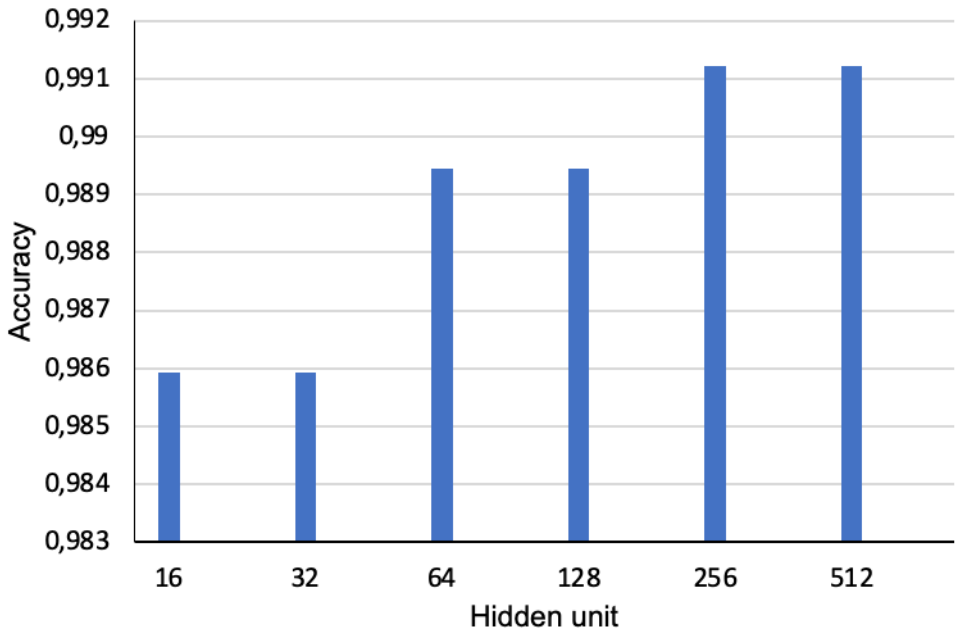


Figure 4.25. Accuracy at different hidden units

As shown in figures (4-24) and (4-25), and also through the remarking of the table (4-23), one can observe the improvement of accuracy as the hidden units increase until a certain point where the accuracy starts to decrease. The interpretation of the accuracy affected by the number of hidden units can be described as follows:

If the number of hidden units is relatively small, the model may not have enough capacity to learn complex patterns in the data. As you increase the number of hidden units, the model's accuracy is expected to improve. This improvement occurs because the network becomes more

capable of capturing important features and relationships in the data, leading to better generalization. When you see an increase in accuracy with a higher number of hidden units in this scenario, it indicates that the initial model was underfitting, and the additional hidden units helped to address this issue by increasing the model's representational power.

If the number of hidden units is already relatively large, adding more hidden units can potentially lead to overfitting. Overfitting occurs when the model becomes too specialized in the training data, and its performance on unseen data starts to degrade. When you see a decrease in validation accuracy (or training accuracy increasing while validation accuracy stagnates or drops) with a higher number of hidden units in this scenario, it indicates that the model is likely overfitting. The additional hidden units allow the model to memorize the training data, but that doesn't necessarily translate to better generalization of new data.[101]

4.2.5 Dropout effects

Figures from (4-26) to (4-28) represent the classification reports and ROC curves at different dropout values (0.35, 0.45, and 0.55)

Table **4**.2**4**. Classification report at dropout 0.35

		precision	recall	F1-score	support
	0	0.9900	0.9900	0.9900	357
	1	0.9900	0.9800	0.9900	212
Accuracy				0.9900	569
Macro avg		0.9900	0.9900	0.9900	569
Weighted avg		0.9900	0.9900	0.9900	569
Accuracy score		0.9894			

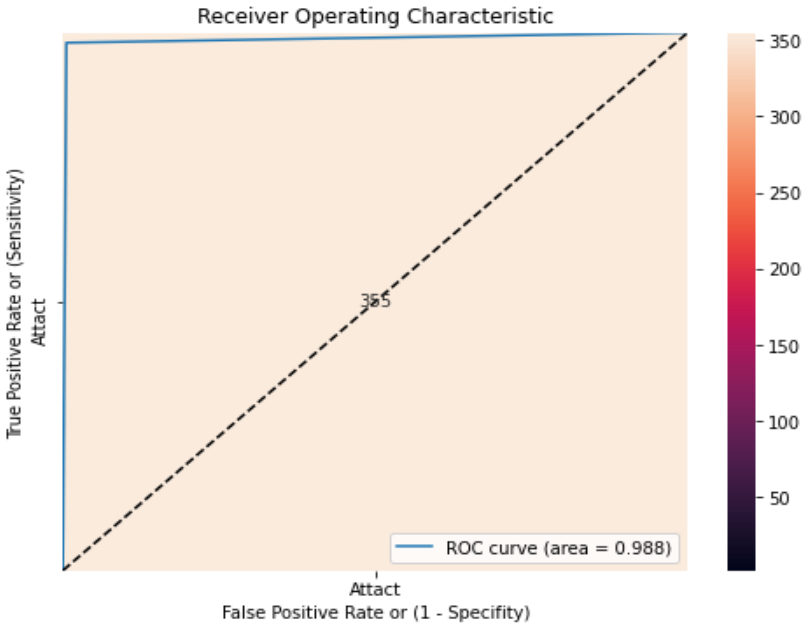


Figure 4.26.The ROC curve for the dropout of 0.35

Table **4**.**25**. Classification report at dropout 0.45

		precision	recall	F1-score	Support
	0	0.9900	0.9900	0.9900	357
	1	0.9900	0.9900	0.9900	212
Accuracy				0.9900	569
Macro avg		0.9900	0.9900	0.9900	569
Weighted avg		0.9900	0.9900	0.9900	569
Accuracy score		0.9912			

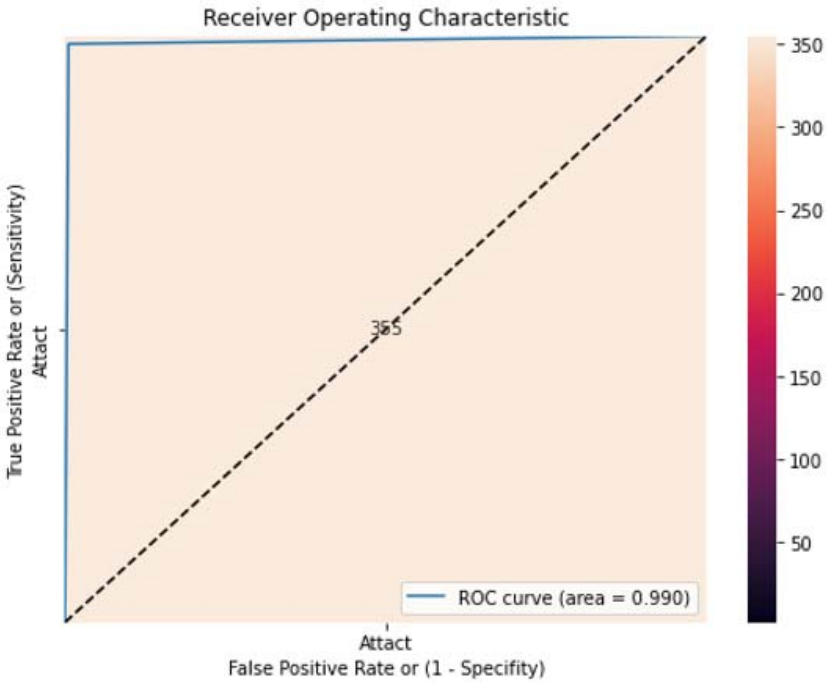


Figure 4.27.The ROC curve for the dropout of 0.45

Table **4**.2**6**. Classification report at dropout 0.55

		Precision	recall	F1-score	support
	0	0.9900	0.9900	0.9900	357
	1	0.9800	0.9800	0.9800	212
Accuracy				0.9800	569
Macro avg		0.9800	0.9800	0.9800	569
Weighted avg		0.9800	0.9800	0.9800	569
Accuracy score		0.9841			

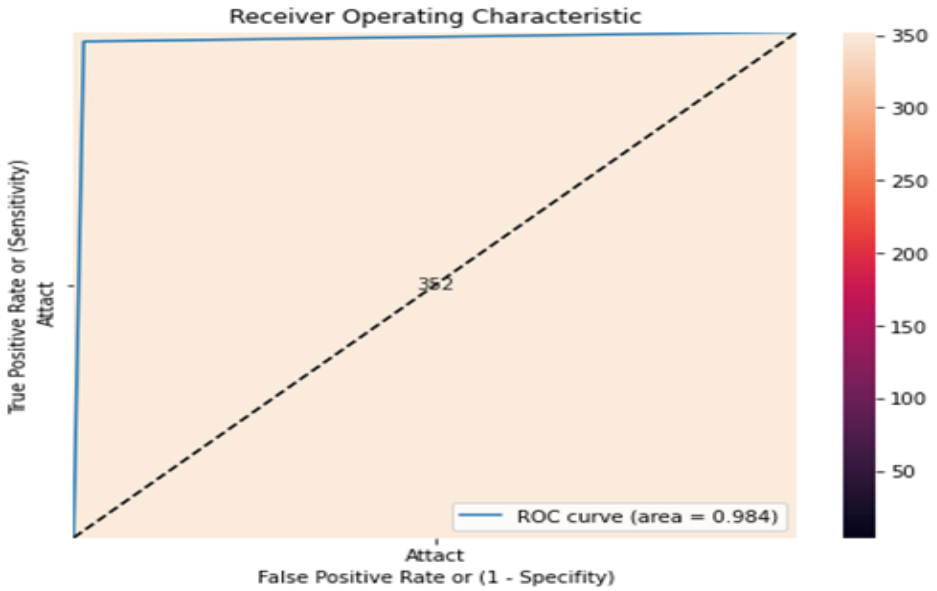


Figure 4.28.The ROC curve for the dropout of 0.55

Table **4.27**.The performance parameters at different dropout

Dropout	Accuracy	Class	Precision	Recall	F1score	AUC
0.35	0.9894	B	0.9900	0.9900	0.9900	0.9880
		M	0.9900	0.9800	0.9900	
0.45	0.9912	B	0.9900	0.9900	0.9900	0.9900
		M	0.9900	0.9900	0.9900	
0.55	0.9841	B	0.9800	0.9800	0.9800	0.9840
		M	0.9800	0.9800	0.9800	

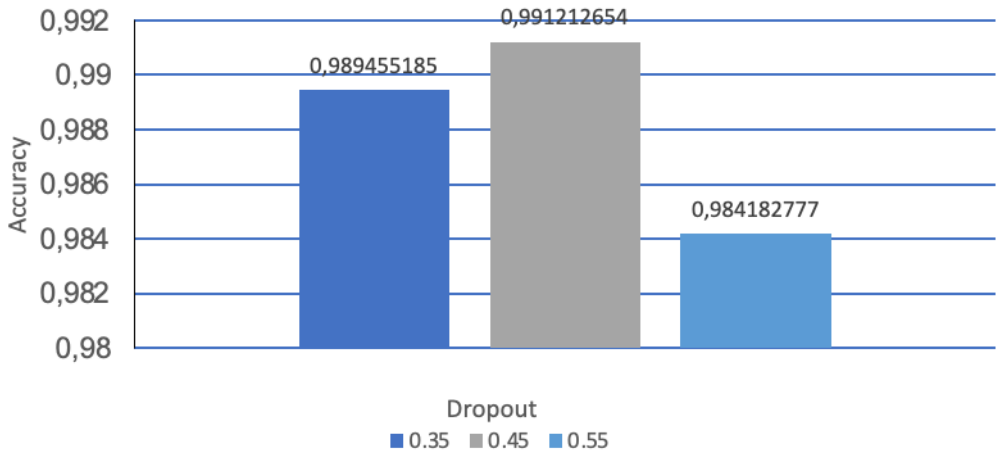


Figure 4.29. Accuracy at different dropout

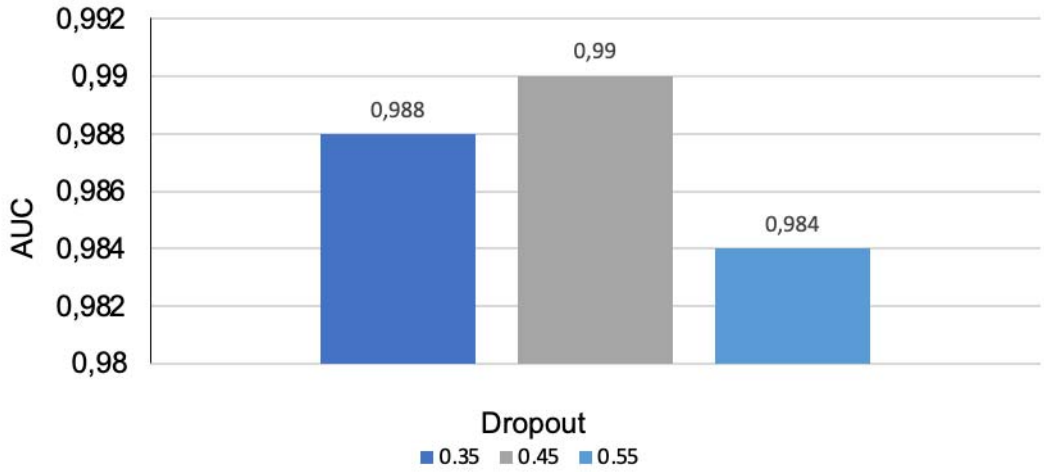


Figure 4.30. The AUC at different dropout rates

As shown in figures (4-29) and (4-30), and as indicated in table (4-27), one can observe the accuracy at different dropout values (0.35, 0.45, and 0.55). The accuracy with the small dropout (0.35) had an accuracy of (0.9894551845342706), this value increased as the dropout increased to 0.45, and then decreased at the dropout of 0.55. The same behavior occurs for AUC values. The interpretation is that higher dropout rates provide stronger regularization, helping prevent overfitting on the training data. As a result, the model may generalize better to new, unseen data, potentially leading to improved accuracy on the validation and test sets. On the other hand, during training, the model's performance on the training set may decrease as dropout rates increase. This is because dropout randomly deactivates neurons, reducing the effective model capacity, as seen with a dropout rate of 0.55. The optimal dropout is then 0.45.

4.2.6 Summary of the final optimal hyperparameters

Table (4-28) presents the final hyperparameter configuration selected for the proposed model after extensive experimental evaluation. These values were chosen based on their superior performance in terms of accuracy, stability, and generalization capability across different training conditions.

Table 4.28. Final optimal hyperparameter values used in the proposed model

Hyperparameter	Final Value
Hidden Units	256
Dropout Rate	0.45
Weight	9
Train–Test Split	90% Training / 10% Testing
Batch Size	32

The combination of these optimized hyperparameters provided the best balance between model complexity, training stability, and resistance to overfitting, forming the final configuration adopted in this study.

4.2.7 Assessment of Hybrid Algorithm Performance on METABRIC Dataset

To realize the proposed hybrid algorithm and to test the ability of its generalization, the hybrid algorithm applied on larger dataset (METABRIC Dataset).

The hybrid algorithm results demonstrated strong capability in both realization and generalization when applied to a larger and more complex dataset.

The model achieved an overall accuracy of 0.9482, indicating consistent classification ability in both categories. For the living category, the model achieved an accuracy of 0.96 and a recall ratio of 0.94, while for the deceased category, it achieved an accuracy of 0.93 and a recall ratio of 0.95.

These values reflect the model's ability to accurately identify positive cases in both categories with minimal bias. Furthermore, the overall and weighted means—both recorded at 0.95—further reinforce the model's strong performance across the entire dataset. These results are closely consistent with the performance of the same hybrid algorithms previously applied to the Wisconsin breast cancer dataset, where similarly high accuracy and balanced metrics were observed.

This consistency highlights the strong generalizability of the proposed hybrid approach, demonstrating that the model maintains its effectiveness when moving from a smaller, simpler dataset, such as Wisconsin, to a larger, more complex dataset, such as Metabrick. This confirms the reliability and scalability of the hybrid methodology used in this study.

Table 4.29. Classification Report

Class	Precision	Recall	F1-Score	Support
Living	0.96	0.94	0.95	1365
Deceased	0.93	0.95	0.94	1144
Accuracy	-	-	0.95	2509
Macro Avg	0.95	0.95	0.95	2509
Weighted Avg	0.95	0.95	0.95	2509

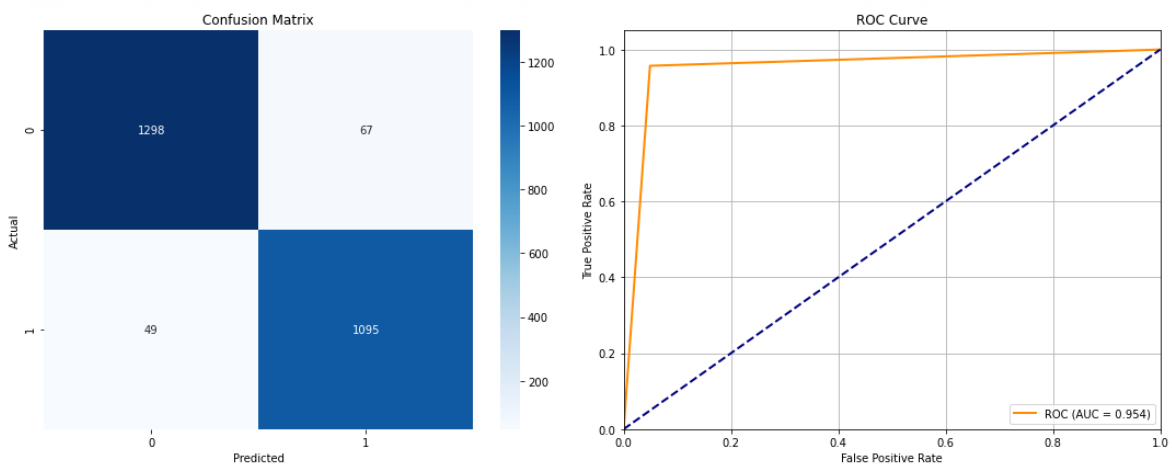


Figure 4.31. Confusion matrix and ROC curve with METABRIC Dataset

The strong results obtained from the METABRIC dataset highlight the effectiveness of the proposed hybrid architecture, which integrates clinical features with high-dimensional genomic information. The confusion matrix shows excellent predictive accuracy, with 1298 true negatives and 1095 true positives, while keeping false classifications low. The balanced sensitivity and specificity (≈95%) indicate that the hybrid model successfully handles the heterogeneous nature of METABRIC, which combines demographic, clinical, and molecular variables.

The ROC curve further emphasizes the model's stability, with an AUC of 0.954 confirming its strong discriminative ability across varying decision thresholds. This demonstrates that the hybrid architecture is capable of learning complex nonlinear relationships that are not easily captured by traditional statistical or shallow machine-learning models. These results validate the ability of the model to generalize beyond smaller datasets such as Wisconsin, and confirm that the METABRIC dataset—being significantly larger and more diverse—benefits substantially from the hybrid learning paradigm. This reinforces the model's suitability for real-world clinical prediction tasks and positions it as a promising candidate for future precision-oncology applications.

4.3. Comparisons with other approaches

The proposed method, which is (Deep MLP, Feature Fused Auto Encoder Neural Network, Weight Tuned Cross-Validated Decision Tree), is compared with the following studies:

1. The study of Md. Milon Islam et al. [4] used five matrices: support vector machine (SVM), KNN, RF, ANN, and LR. Both of these studies used the (WDBC) dataset. Table (4-30) and Figure (4-31) show the comparison between them. The proposed method exhibited enhanced performance parameters.

Table 4.30. Performance parameters for different matrices

Algorithm	Accuracy	Recall	F1score	Precession
SVM	0.9741	1.0000	0.9777	0.9565
KNN	0.9741	0.9782	0.9782	0.9782
RF	0.9571	0.9565	0.9670	0.9777
ANN	0.9857	1.0000	0.9890	0.9782
LR	0.9571	0.9574	0.9677	0.9782
Proposed	**0.9947**	**0.9900**	**0.9900**	**0.9900**

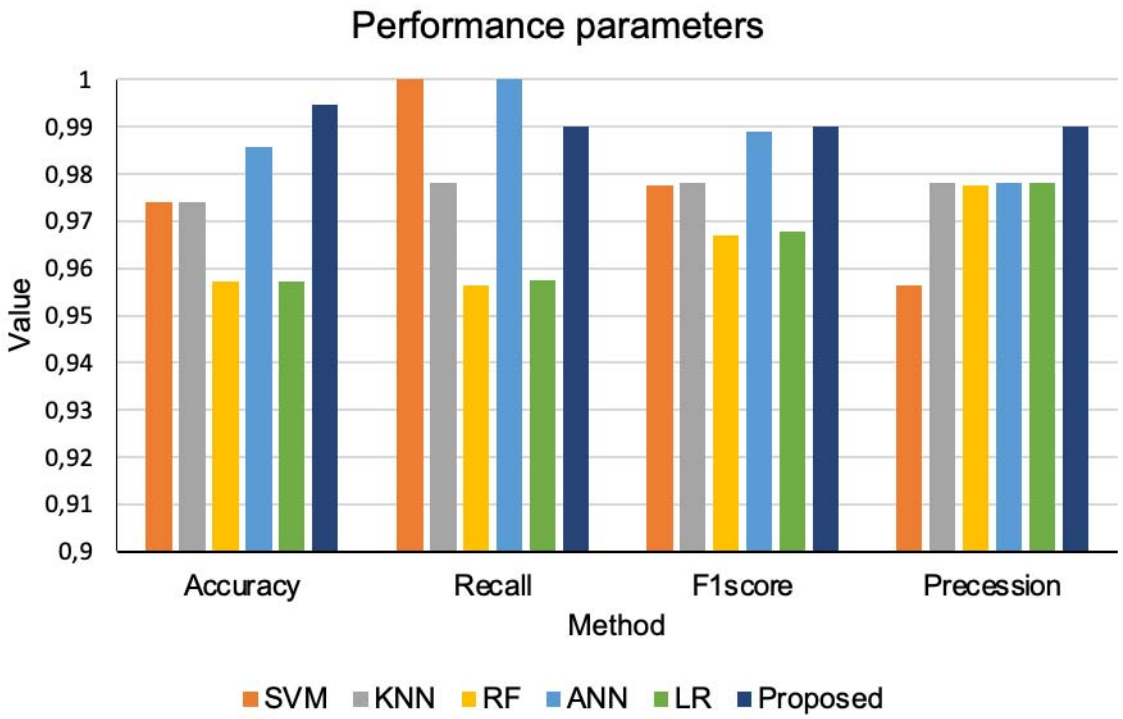


Figure 4.31. Performance parameters of different matrices with the proposed approach

2. The proposed method, which is (Deep MLP, Feature Fused Auto Encoder Neural Network, Weight Tuned Cross-Validated Decision Tree), is compared with the study of Nirdosh Kumar [64], which used (LR, SVM, KNN, and Naïve Bayes) algorithms. For the WDBC dataset, the performance parameters are shown in the table (4-31) and the figure (4-32). The proposed work achieves higher accuracy and recall than LR, KNN, and Naïve Bayes, while matching the performance of SVM.

Table 4.31. Performance parameters for different matrices

Algorithm	Accuracy	Recall
LR	0.9649	0.9565
SVM	0.9824	0.9900
KNN	0.9720	0.9808
Naïve Bayes	0.9474	0.9545
proposed	**0.9947**	**0.9900**

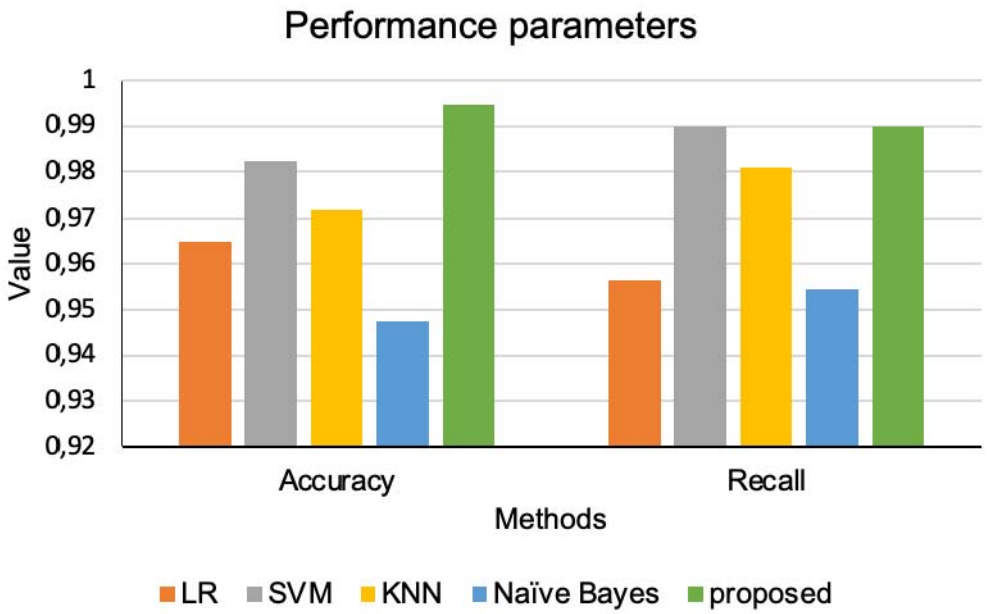


Figure 4.32. Performance parameters of different matrices with the proposed approach

3. The performance parameters of the study of G. Battineni et al [102] and the proposed work are shown in Table (4-32) and Figure (4-33). The proposed work exhibited higher values of accuracy, recall, and AUC than SVM, LR, and KNN.

Table **4.32**.Performance parameters for different matrices

Algorithm	Accuracy	Recall	precision	AUC
SVM	0.9766	0.9375	1.0000	0.9361
LR	0.9766	0.9531	0.9838	0.9123
KNN	0.9649	0.9062	1.0000	0.9371
proposed	**0.9947**	**0.9900**	**0.9900**	**0.9930**

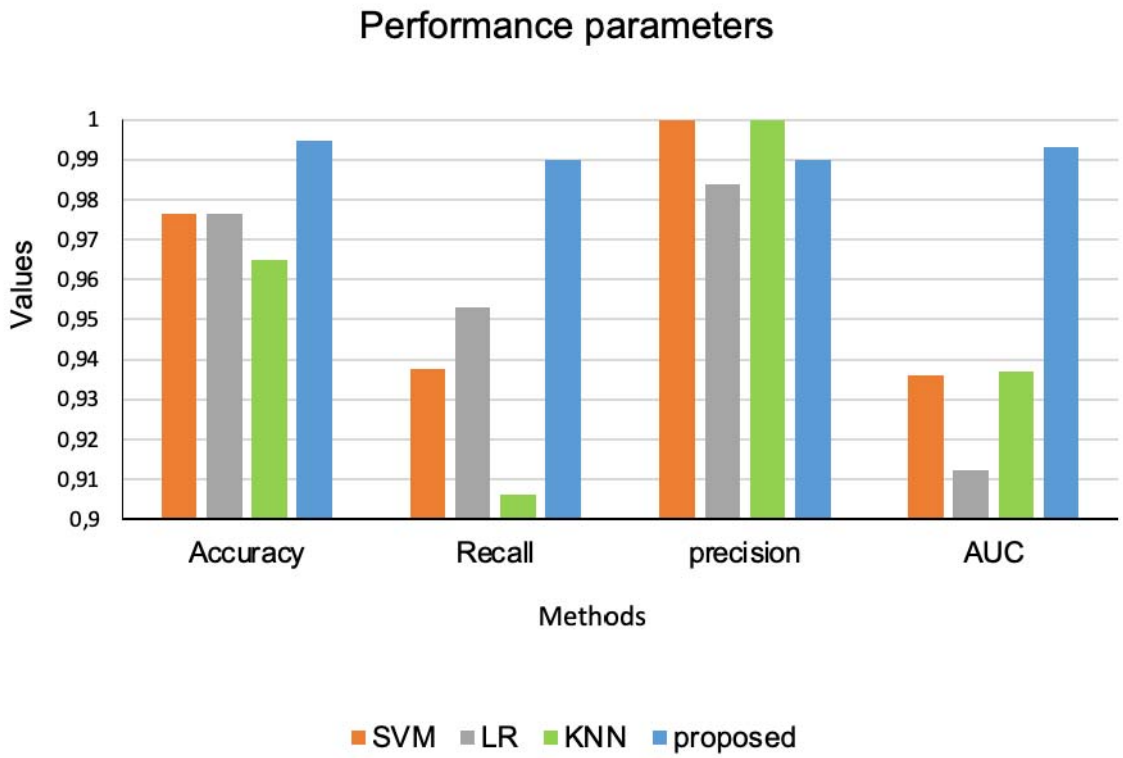


Figure **4.33**.Performance parameters of different matrices with the proposed approach

4. The study of P. Karatza, et al [66], which uses RF, NN, and ENN in the diagnosis of breast cancer from the WDBC dataset . This study is compared with the proposed work. All performance parameters of the proposed work were larger. Table (4-33) and Figure (4-34) show this comparison.

Table **4.33**. Performance parameters for different matrices

Algorithms	Accuracy	Recall	AUC
RF	0.9649	0.9437	0.9600
NN	0.9410	0.8836	0.9300
ENN	0.9660	0.9494	0.9600
proposed	**0.9947**	**0.9900**	**0.9900**

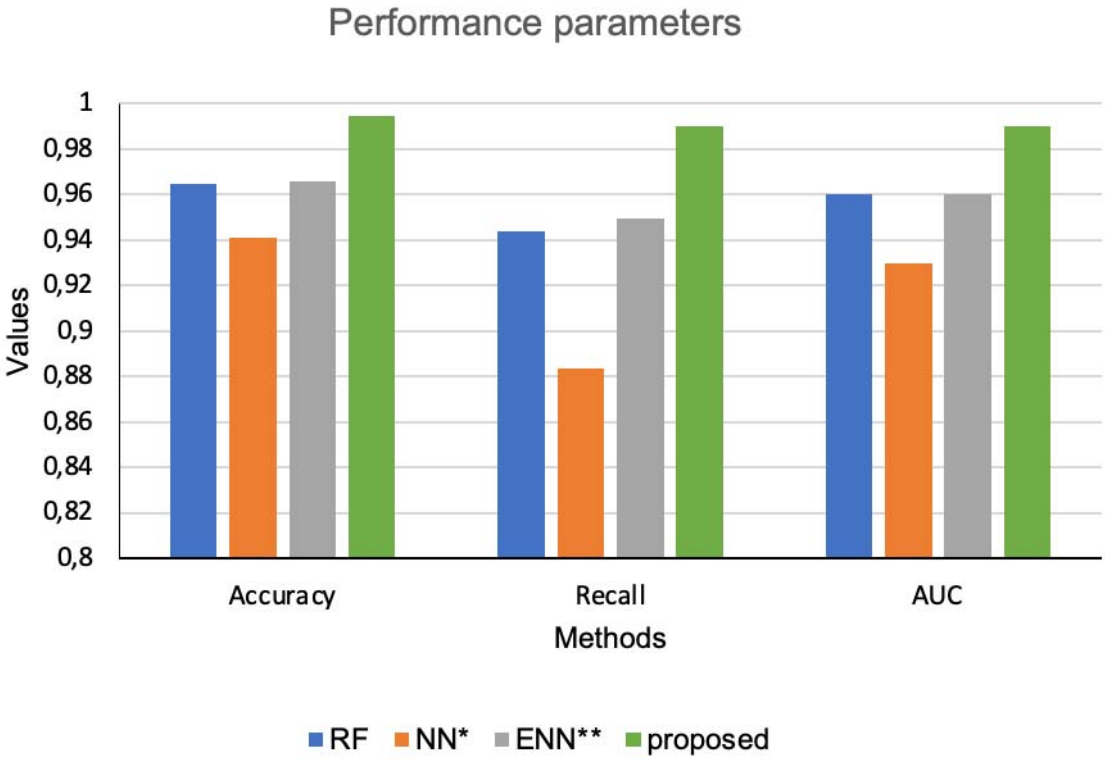


Figure **4.34**.Performance parameters of different matrices with the proposed approach

5. Table (4-34) and Figure (4-35) show the performance parameters of the study presented by Jiande Wu, and Chindo Hicks. This study evaluated four classification models —Support Vector Machines, K-nearest neighbors, Naïve Bayes, and Decision Trees —using features selected for Breast Cancer Type Classification Using Machine Learning. SVM achieved the highest Accuracy and Recall of 90% and 87%, respectively, while DT achieved the highest Specificity of 91%.

Table **4.3**4. Performance parameters for different matrices

Algorithms	Accuracy	Recall	Specificity
KNN	0.8700	0.7600	0.8800
NGB	0.8500	0.6800	0.8700
DT	0.8700	0.5400	0.9100
SVM	0.9000	0.8700	0.9000
proposed	**0.9947**	**0.9900**	**0.9900**

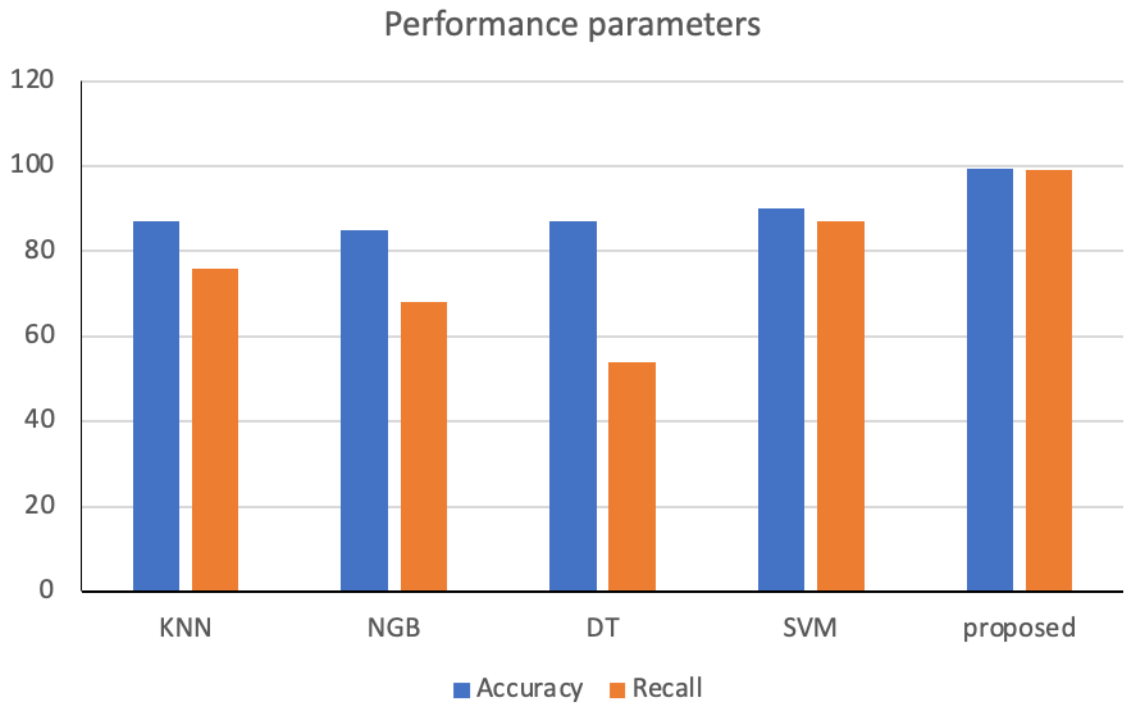


Figure 4.35.Performance parameters of different matrices with the proposed approach

6. Table (4-35) and Figure (4-36) show the performance parameters of the study presented by Vu Pham Thao Vy et al., This study aimed to develop a machine-learning classification model to differentiate ductal carcinoma in situ and minimally invasive breast cancer using clinical characteristics, mammography findings, ultrasound findings, and histopathology features. The model in this study identified the five most important features as calcifications on mammograms, lymph node presence, microcalcifications on histopathology, the shape of the mass on ultrasound, and the orientation of the mass on ultrasound.

Table **4.3**5. Performance parameters for different matrices

Model	Accuracy	F1 Score	Recall	Precision
XGBoost	0.8400	0.8700	0.9100	0.8200
GaussianNB	0.7500	0.7900	0.8800	0.7200
KNeighbors	0.6300	0.7300	0.8700	0.6200
DecisionTree	0.7300	0.7600	0.7700	0.7500
RandomForest	0.8200	0.8400	0.8900	0.8100
Proposed	**0.9947**	**0.9900**	**0.9900**	**0.9900**

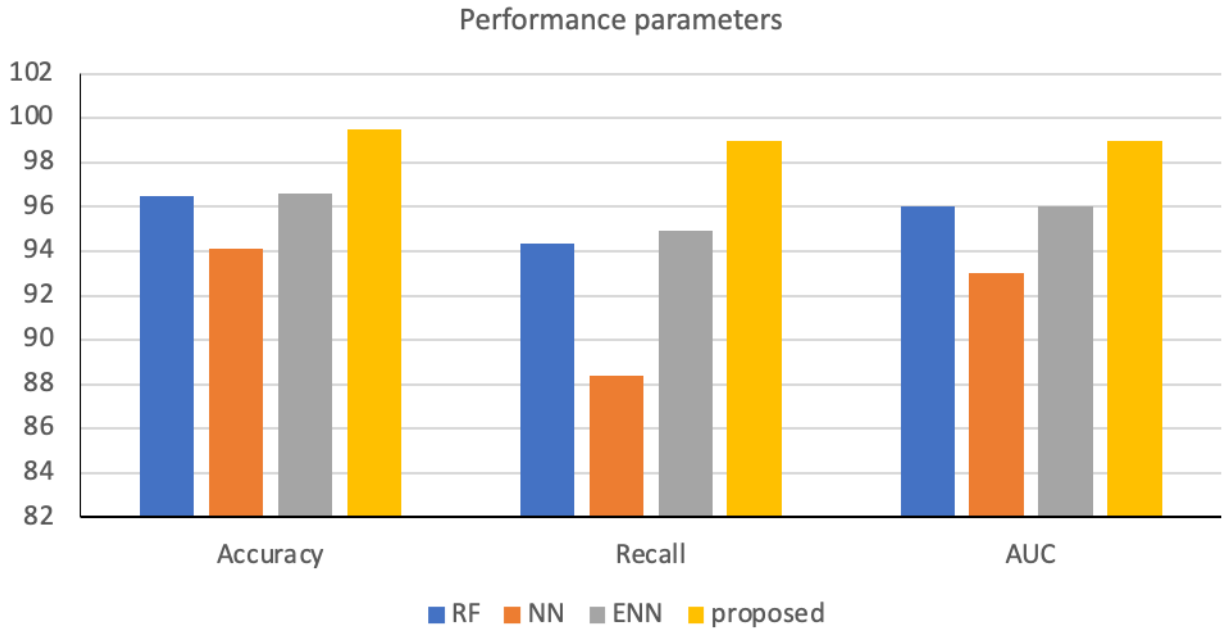


Figure 4.36. Performance parameters of different matrices with the proposed approach

7. Table (4-36) and Figure (4-37) show the performance parameters of the study presented by Mohammed Amine Naji et al. This study used SVM, RF, LR, DT, and KNN for breast cancer prediction and diagnosis. They studied accuracy, precision, sensitivity, and F-Measure. SVM achieved the highest accuracy ≈ of 97.2% on the WBCD dataset.

Table 4.36. Performance parameters for different matrices

Algorithms	Accuracy	Precision	Sensitivity	F-Measure	Class
SVM	0.984	0.9800	0.9400	0.9600	Benign
		0.9700	0.9900	0.9800	Malignant
R F	0.998	0.9600	0.9400	0.9500	Benign
		0.9700	0.9800	0.9700	Malignant
L R	0.955	0.9800	0.9100	0.9400	Benign
		0.9500	0.9900	0.9700	Malignant
D T	0.988	0.9400	0.9200	0.9300	Benign
		0.9600	0.9700	0.9600	Malignant
KNN	0.946	0.9200	0.9100	0.9100	Benign
		0.9500	0.9600	0.9500	Malignant
proposed	**0.9947**	**0.9900**	**1.0000**	**1.0000**	Benign
		1.0000	**0.9900**	**0.9900**	Malignant

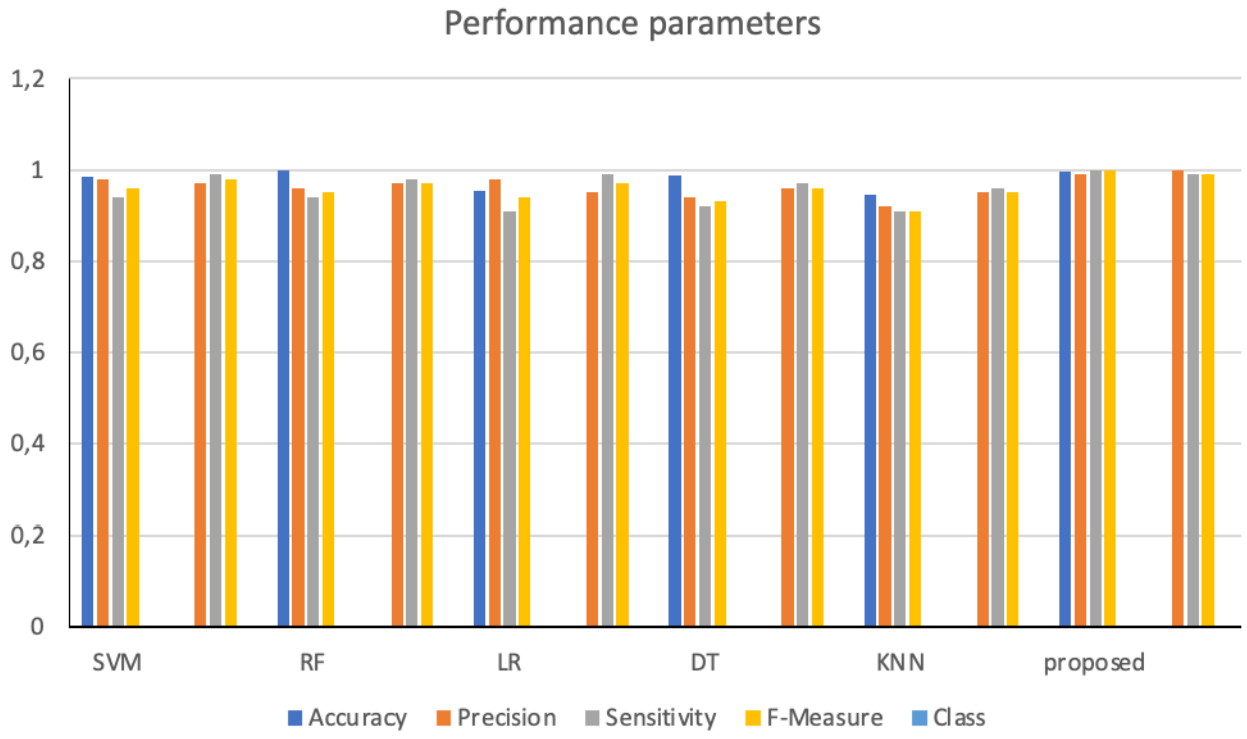


Figure 4.37. Performance parameters of different matrices with the proposed approach

8. Table (4-37) and Figure (4-38) show the performance parameters of the study presented by A. I. Aishwarja et al [103] and also those of the proposed work. The proposed work has the highest value of all parameters.

Table 4.37. Performance parameters for different matrices

Algorithms	Accuracy	precision	Recall
RF	0.9590	0.9720	0.9630
KNN	0.9590	0.9600	0.9850
SVM	0.9720	0.9730	0.9900
Naïve Bayes	0.9240	0.9260	0.9550
proposed	**0.9940**	**0.9900**	**0.9900**

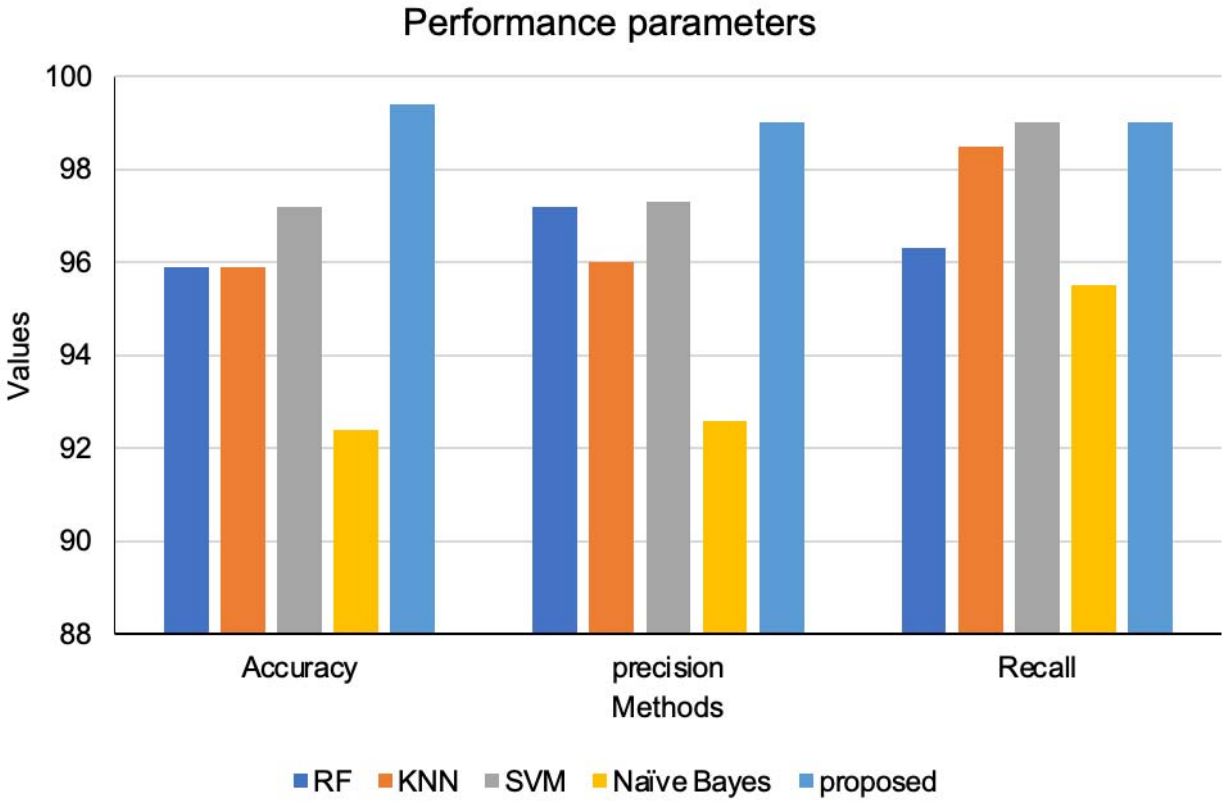


Figure 4.38. Performance parameters of different matrices with the proposed approach

9. The proposed work is compared with the work of S. Sakib et al [71] that used SVM, DT, LR, RF, KNN, and NN with the WDBC dataset. The accuracy, recall, and F1-score were largest for the proposed work as shown in the table (4-38) and Figure (4-39).

Table 4.38. Performance parameters for different matrices

Algorithms	Accuracy	Recall	Precision	F1 score
SVM	0.9473	0.8570	0.9470	0.90000
DT	0.920	0.8810	0.9490	0.9140
LR	0.9508	0.8570	0.9730	0.9110
RF	0.9666	0.9290	1.0000	0.9630
KNN	0.9332	0.7860	1.0000	0.8800
NN	0.9035	0.7620	0.9700	0.8540
proposed	**0.9947**	**0.9900**	**0.9900**	**0.9900**

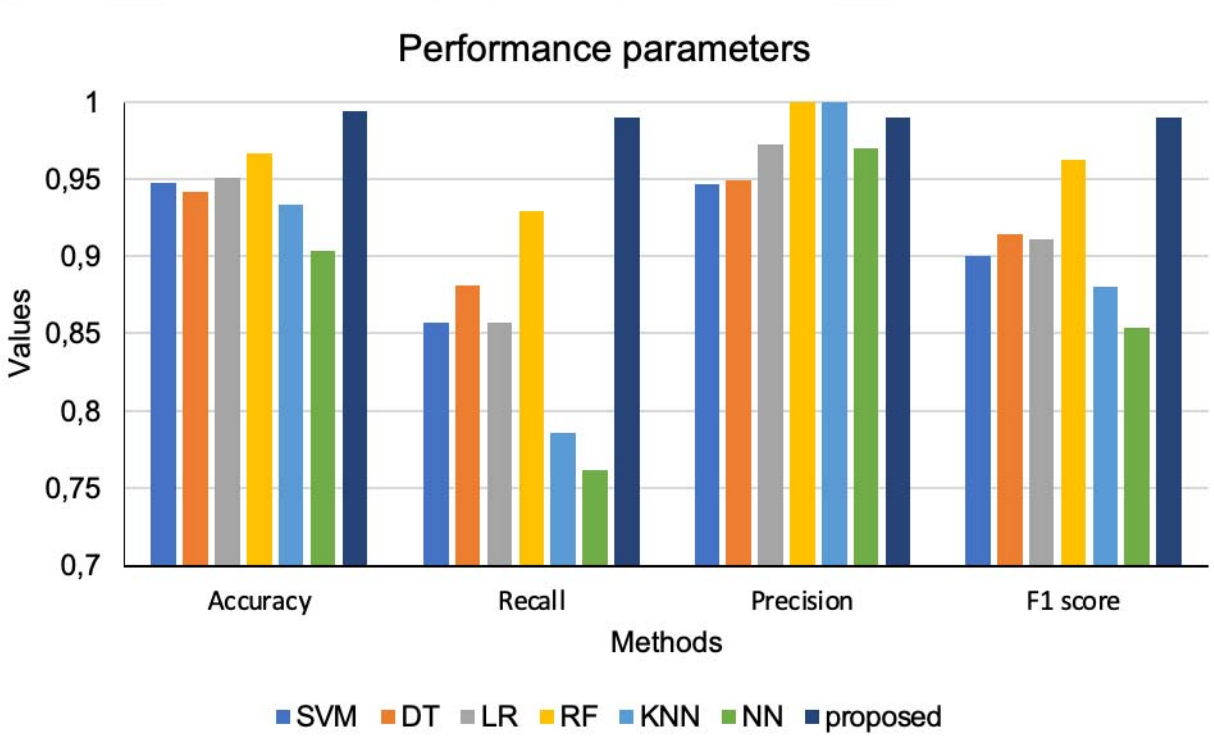


Figure 4.39. Performance parameters of different matrices with the proposed approach

10. The proposed work is compared with the work of (Bharti Thakur et al. 2022) [11]. Machine learning classifiers, including SVM, Decision Tree, AdaBoost, random forests, and ANN, were used in this paper to detect breast cancer. SVM gives 86.9% accuracy; decision tree 85.3%; random forest 86.3%; AdaBoost 68.5%; and ANN 87.43%, as shown in the table (4-39) and Figure (4-40).

Table 4.39. Accuracy for different algorithms at different train/ test splitting

Algorithm	Accuracy 70% / 30%	Accuracy 80% / 20%	Accuracy 90% / 10%
D T	82%	83%	86%
AdaBoost Classifier	66%	67%	68%
R F	82%	84%	85%
SVM	81%	83%	84%
ANN	79%	80%	82%
proposed	**97%**	**98%**	**99%**

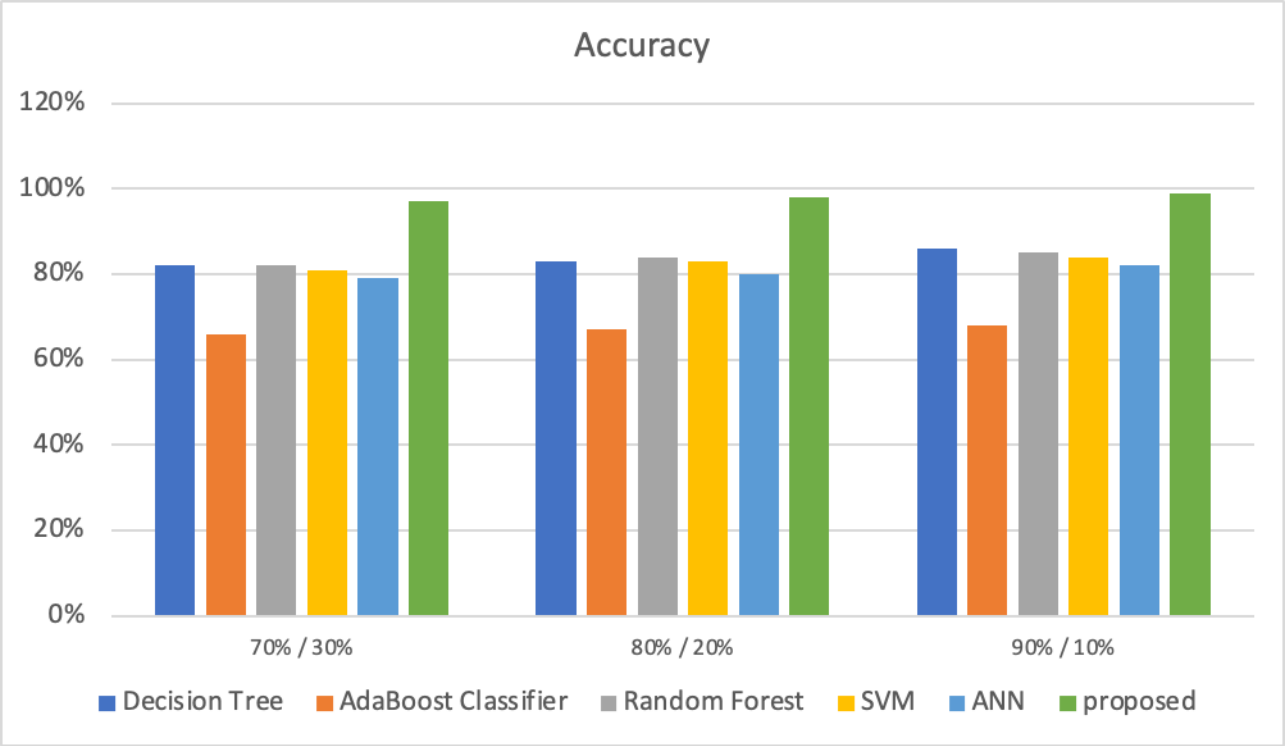


Figure 4.40. Accuracy for different algorithms at different train/ test splitting

11. The proposed work is compared with the work of M. Monirujjaman Khan et al [73], which used RF, LAR, DT, and KNN algorithms with the WDBC dataset. The accuracy was the largest for the proposed work, as shown in Table (4-40) and Figure (4-41).

Table 4.40. Performance parameters for different matrices

Algorithm	Accuracy	Class	F1-score
RF	0.9561	B	0.9400
		M	0.9700
LR	0.9860	B	0.9700
		M	0.9800
DT	0.9473	B	0.9300
		M	0.9600
KNN	0.9035	B	0.8500
		M	0.9300
Proposed	**0.9947**	B	**1.0000**
		M	**0.9900**

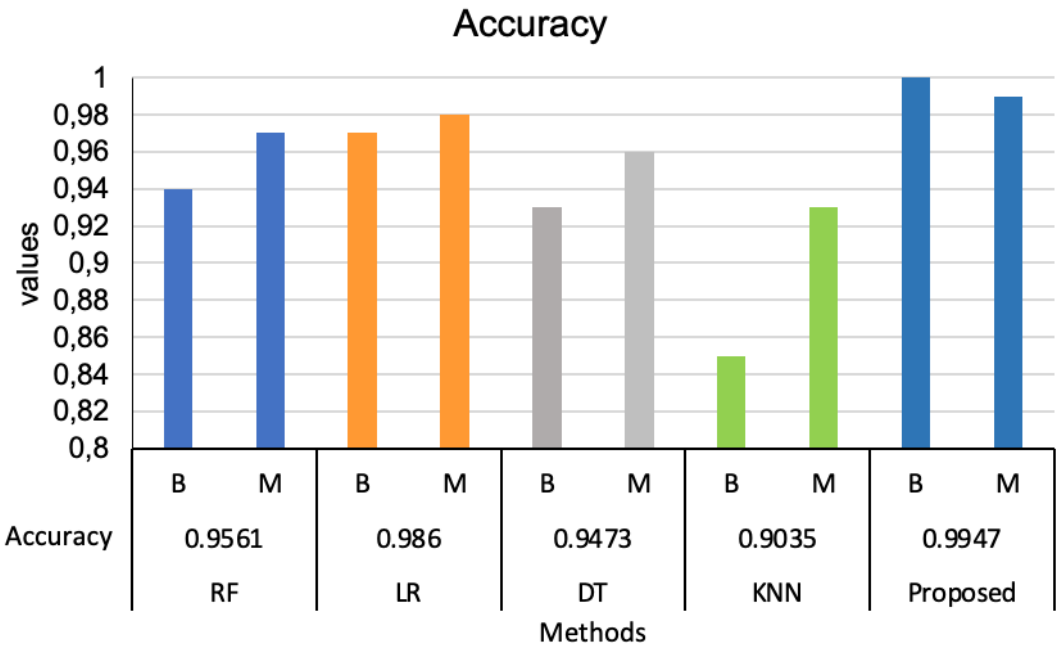


Figure 4.41. Accuracy of different matrices with the proposed approach

12. The proposed work is compared with the work of H. Prajapati [104], which used LR, RF, and DT with the WDBC dataset. The accuracy, precision, and recall were the largest for the proposed work, as shown in the table (4-41) and figure (4-42).

Table 4.41. Performance parameters for different matrices

Algorithms	Accuracy	Precision	Recall
LR	0.9640	0.9700	0.9600
RF	0.9730	0.9800	0.9700
DT	0.9380	0.93500	0.9350
Proposed	**0.9947**	**0.9900**	**0.9900**

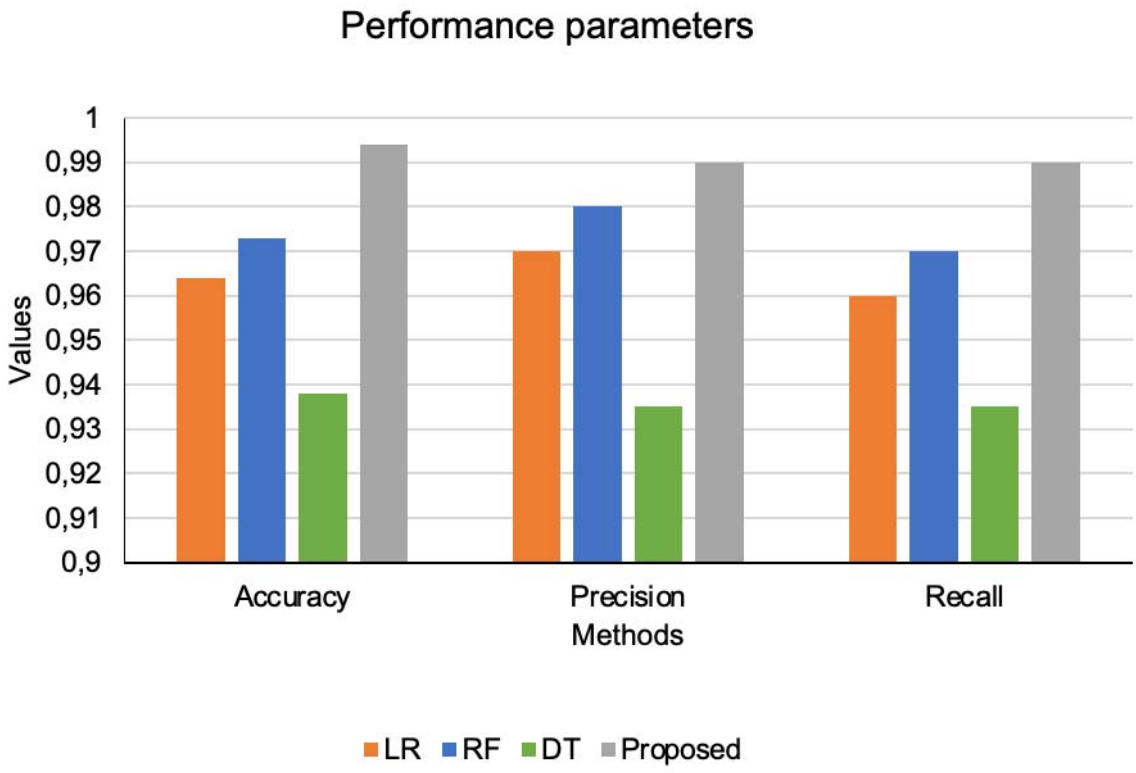


Figure 4.42.Performance parameters of different matrices with the proposed approach

13. The performance parameters of the study presented by V. Nemade and V. Fegade [74] and the proposed work are shown in Table (4-42) and Figures (4-43) and (4-44). The proposed work exhibited higher values of accuracy, precision, recall, and F1-score than KNN, SVM, DT, Naïve Bayes, and LR.

Table 4.42. Performance parameters for different matrices

Algorithm	Accuracy	Class	Precision	Recall	F1-score
KNN	**0.9600**	B	0.9300	1.0000	0.9600
		M	1.0000	0.8900	0.9400
SVM	**0.9500**	B	0.9700	0.9400	0.9500
		M	0.9200	0.9600	0.9400
DT	**0.9700**	B	0.9700	0.9900	0.9800
		M	0.9800	0.9600	0.9700
Naïve Bayes	**0.9000**	B	0.9200	0.9100	0.9200
		M	0.8800	0.8900	0.8800
LR	**0.9600**	B	0.9700	0.9700	0.9700
		M	0.9600	0.9600	0.9600
proposed	**0.9900**	B	0.9900	1.0000	1.0000
		M	1.0000	0.9900	0.9900

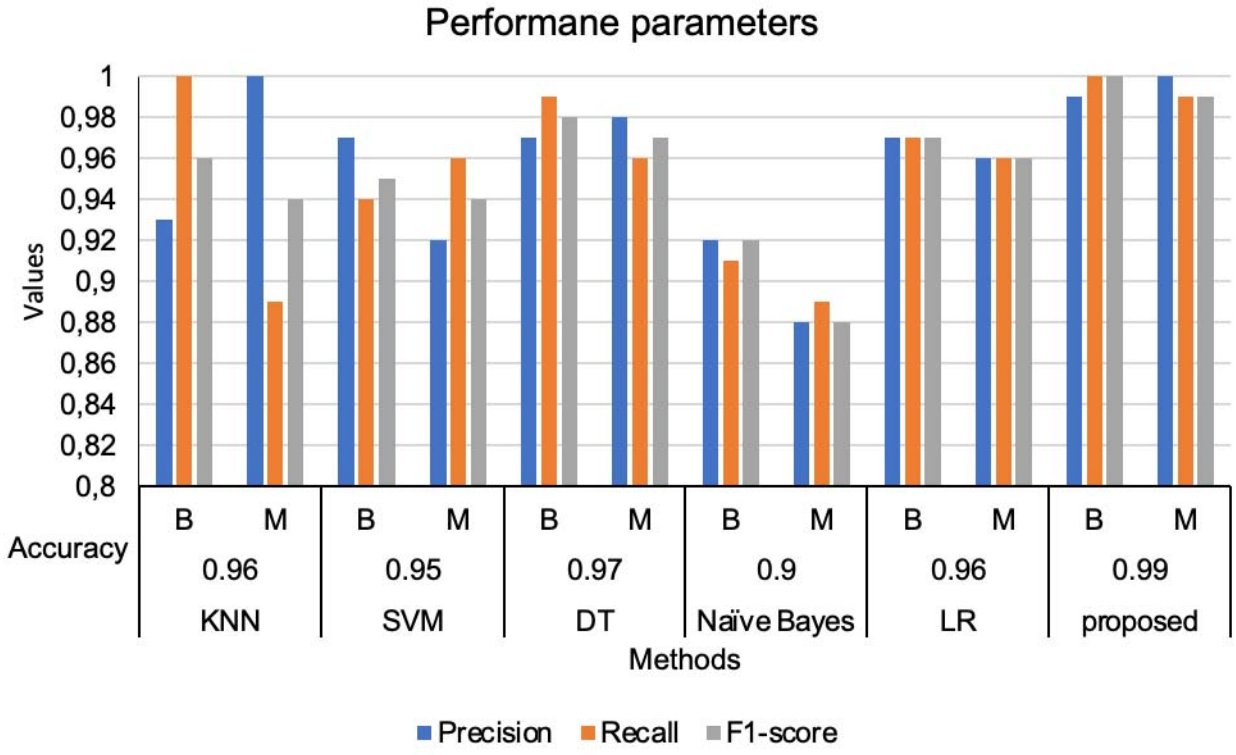


Figure 4.43.Performance parameters of different matrices with the proposed approach

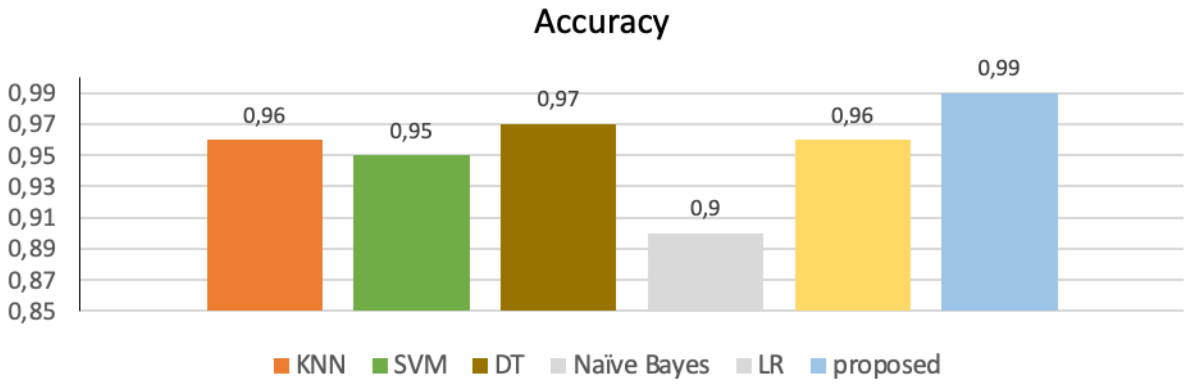


Figure 4.44. Accuracy of different matrices with the proposed approach

14. The proposed work is compared with the work of (Khandaker Mohammad Mohi Uddin et al. 2023. In this research, the Wisconsin Breast Cancer Dataset (WBCD) from the UCI machine learning repository has been used as a training set to compare the performance of various machine learning techniques. Different kinds of machine learning classifiers such as support vector machine (SVM), Random Forest (RF), K-nearest neighbors(K-NN), Decision tree (DT), Naïve Bayes (NB), Logistic Regression (LR), AdaBoost (AB), Gradient Boosting (GB), Multi-layer perceptron (MLP), Nearest Cluster Classifier (NCC), and voting classifier (VC) have been used for comparing and analyzing breast cancer into benign and malignant tumors. Various metrics, such as error rate, Accuracy, Precision, F1-score, and recall, have been used to evaluate the model's performance. Each Algorithm's accuracy has been assessed to determine the best one. Based on the analysis, the Voting classifier has the highest accuracy (98.77%) and the lowest error rate. , as shown in the table (4-43) and figure (4-45).

Table 4.43. Performance parameters for different matrices

Methods	Precision	Recall	F-Measure	Accuracy
SVM	0.9828	0.9761	0.9792	0.9807
RF	0.9391	0.9365	0.9378	0.9420
KNN	0.9722	0.9604	0.9658	0.9684
DT	0.9399	0.9356	0.9377	0.9420
NB	0.9073	0.8998	0.9033	0.9104
LR	0.9855	0.9807	0.9830	0.9842
AB	0.9627	0.9581	0.9603	0.9631
GB	0.9557	0.9539	0.9548	0.9578
MLP	0.9755	0.9718	0.9736	0.9754
NCC	0.9344	0.09186	0.9254	0.9315
VC (LR + SVM)	0.9883	0.9854	0.9868	0.9877
Proposed	**0.9900**	**0.9900**	**0.9900**	**0.9947**

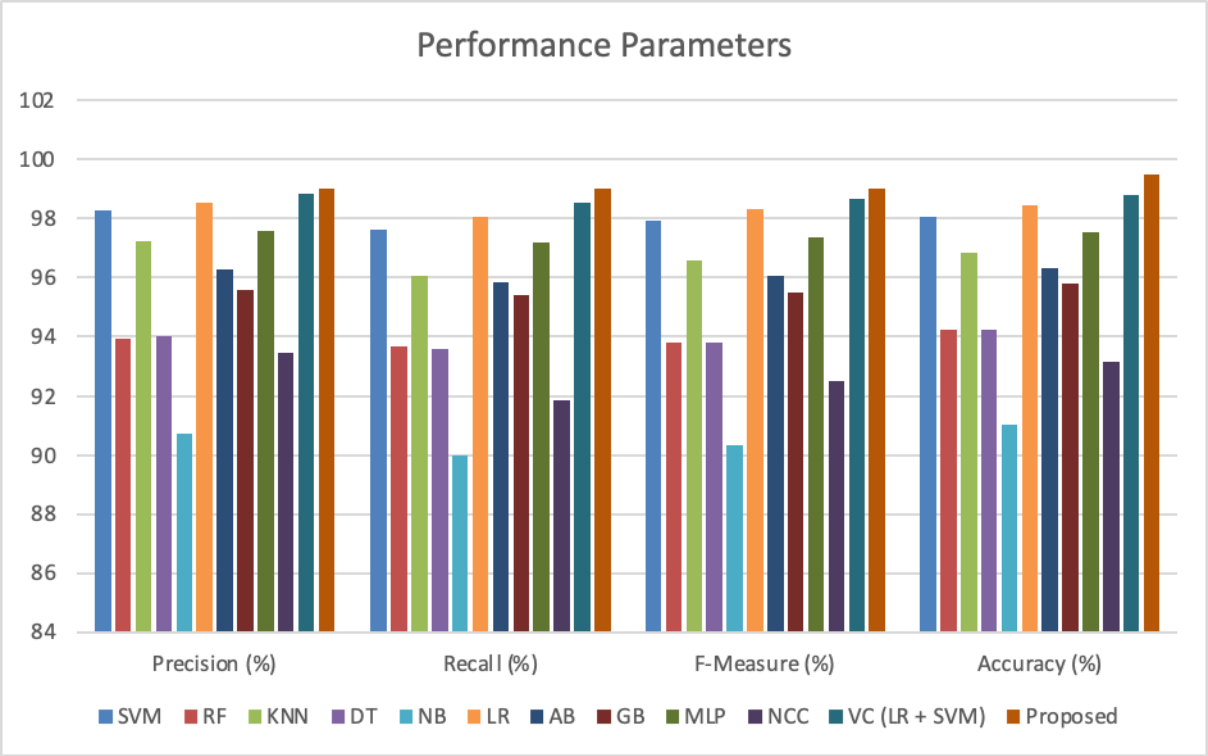


Figure 4.45. Performance parameters of different matrices with the proposed approach

15. The proposed work is compared with the work of **Dinesh & Kalyanasundaram (2023), which** compared SVM, KNN, Decision Tree, and Random Forest using UCI's breast cancer data. Random Forest outperformed the others with ~95% accuracy, compared with 92% (SVM), 90% (KNN), and 88% (Decision Tree), as shown in the table (4-44) and figure (4-46).

Table 4.44. Performance parameters for different matrices

Algorithm	**Accuracy**
Random Forest	0.9500
SVM	0.9200
KNN	0.9000
Decision Tree	0.8800
Proposed	**0.9947**

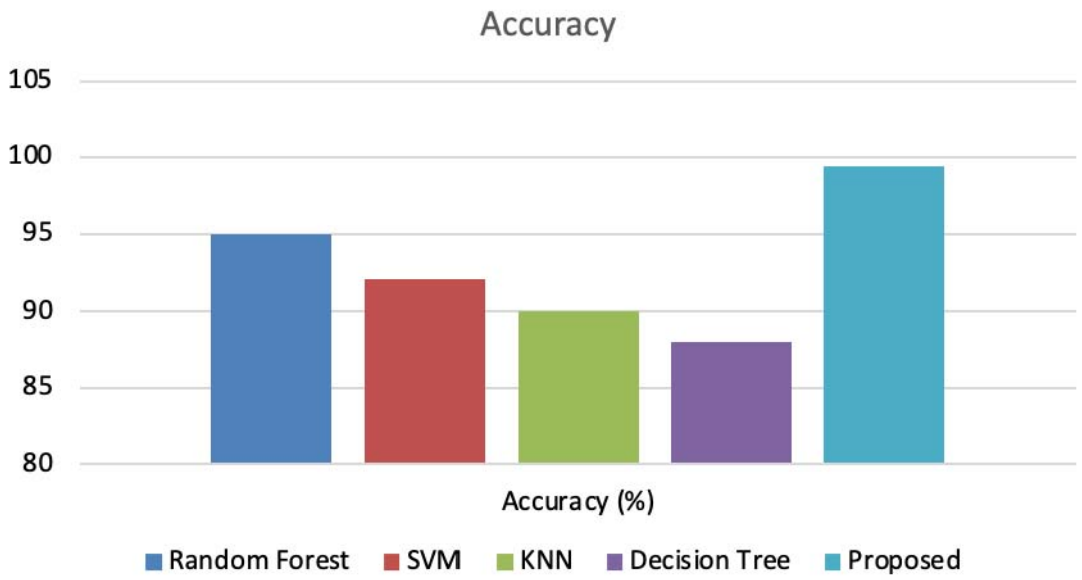


Figure 4.46. Accuracy at different matrices with the proposed approach

16. The proposed work is compared with the work of Aiman Fatima et al. 2023). In this study, different algorithms are examined. SVC obtained 86.36%, Decision Tree 86.18%, and Random Forest 86.00%. The deep learning model, more precisely, a neural network, outperformed these results with a highest 93% accuracy, as shown in the table (4-45) and figure (4-47).

Table 4.45. Performance parameters for different matrices

Classifier	**Accuracy**	**Precision**	**Recall**	**F1- Score**
D T	0.8618	0.7800	0.7700	0.7700
R F	0.8690	0.7900	0.7700	0.7800
SVC	0.8836	0.8500	0.7500	0.7800
N N	0.9300	0.9800	0.8700	0.9200
Proposed	**0.9947**	**0.9900**	**0.9900**	**0.9900**

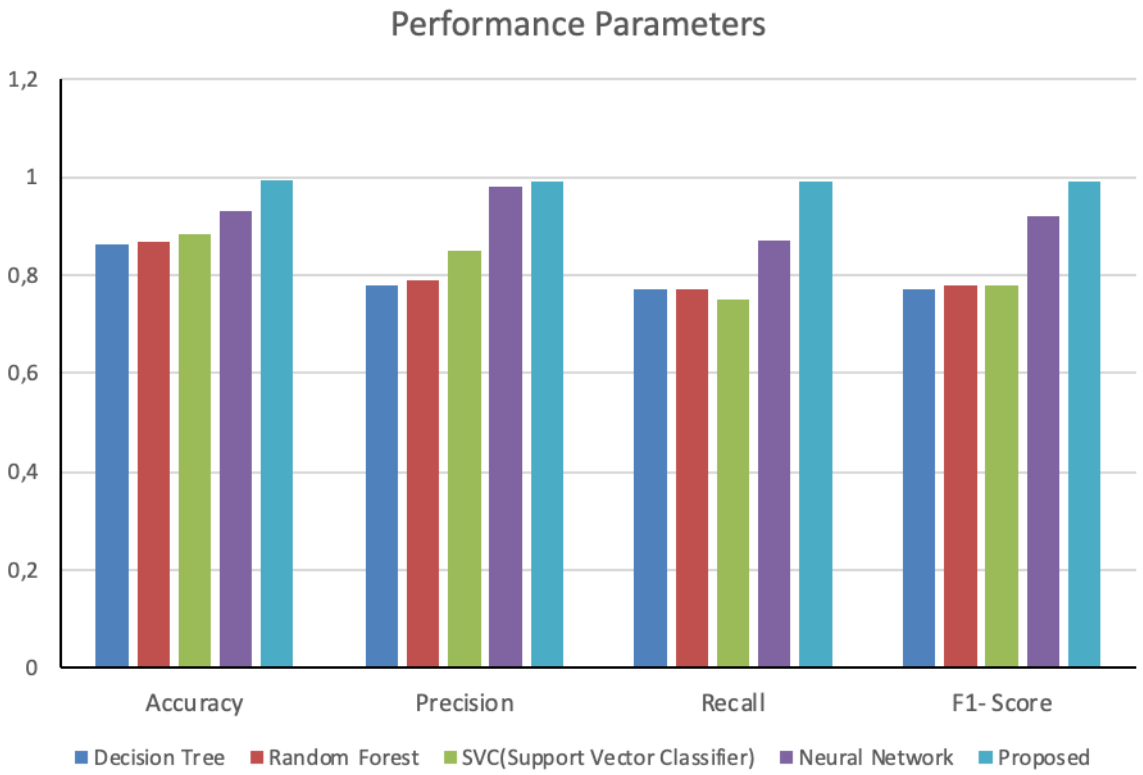


Figure 4.47. Accuracy of different matrices with the proposed approach

17. The proposed work is compared with the work of Ruiming Zeng et al. 2024 In this work, an efficient and robust deep learning framework called FastLeakyResNet-CIR, an improved ResNet architecture, is proposed for breast cancer detection and classification, as shown in the table (4-46) and figure (4-48).

Table 4.46.Performance parameters for different algorithms

Model	Accuracy	Precision	Recall	F1-Score
FastLeakyResNet-CIR	0.9894	0.9943	0.9836	0.9889
ResNet50	0.9263	0.9483	0.8969	0.9219
InceptionV3	0.9123	0.9107	0.9082	0.9094
ResNet18	0.8614	0.8763	0.8314	0.8533
VGG16	0.7945	0.7855	0.7925	0.7990
Proposed	**0.9947**	**0.9900**	**0.9900**	**0.9900**

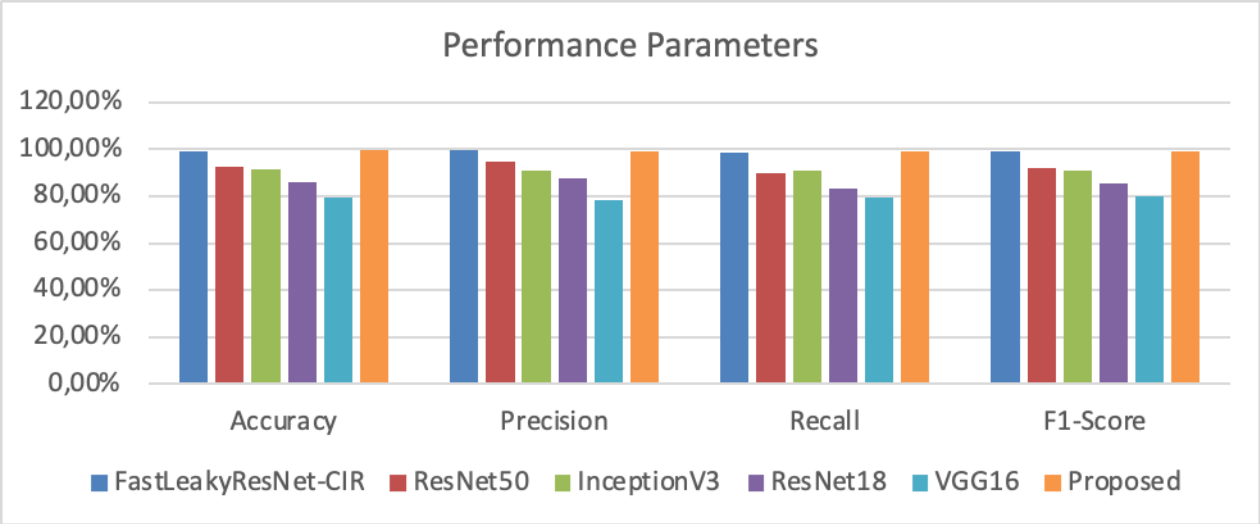


Figure 4.48. Accuracy at different matrices with the proposed approach

18. The proposed work is compared with the work of (Patnala S. R. et al. 202 The objective of this investigation was to improve the diagnosis of breast cancer by combining two significant datasets: the Wisconsin Breast Cancer Database and the DDSM Curated Breast Imaging Subset (CBIS-DDSM). A hybrid deep learning algorithm consisting of CNNs and stochastic gradients is used for the detection of breast cancer, utilizing an accuracy is 96% precision of 95%, a Recall is 95%, and an AUC-ROC value of 0.96, as shown in the table (4-47) and figure (4-49).

Table 4.47. Performance parameters for different algorithms

Model	Accuracy	Sensitivity	Specificity	Precision	Recall	AUC
CNN	0.8700	0.86	0.8500	0.8700	0.8600	0.8700
CNN-AlexNet	0.8900	0.85	0.8600	0.8800	0.8900	0.8900
CNN-VGG	0.9000	0.91	0.8900	0.9100	0.9200	0.9000
CNN-NiN	0.9100	0.90	0.0900	0.9100	0.9000	0.9100
CNN-GoogleNet	0.9100	0.90	0.8900	0.9000	0.9100	0.9100
CNN-ResNet	0.9300	0.90	0.9200	0.9200	0.9300	0.9300
CNN-DenseNet	0.9200	0.93	0.9000	0.9000	0.9100	0.9200
CNN-VGG + SGD	0.9600	0.95	0.9600	0.9500	0.9600	0.9600
CNN-GoogleNet + SGD	0.9600	0.95	0.9500	0.9500	0.9600	0.9600
CNN-DenseNet + SGD	0.9600	0.95	0.9500	0.9600	0.9500	0.9600
Proposed	**0.9947**	**0.9900**	**0.9900**	**0.9900**	**0.9900**	**0.9900**

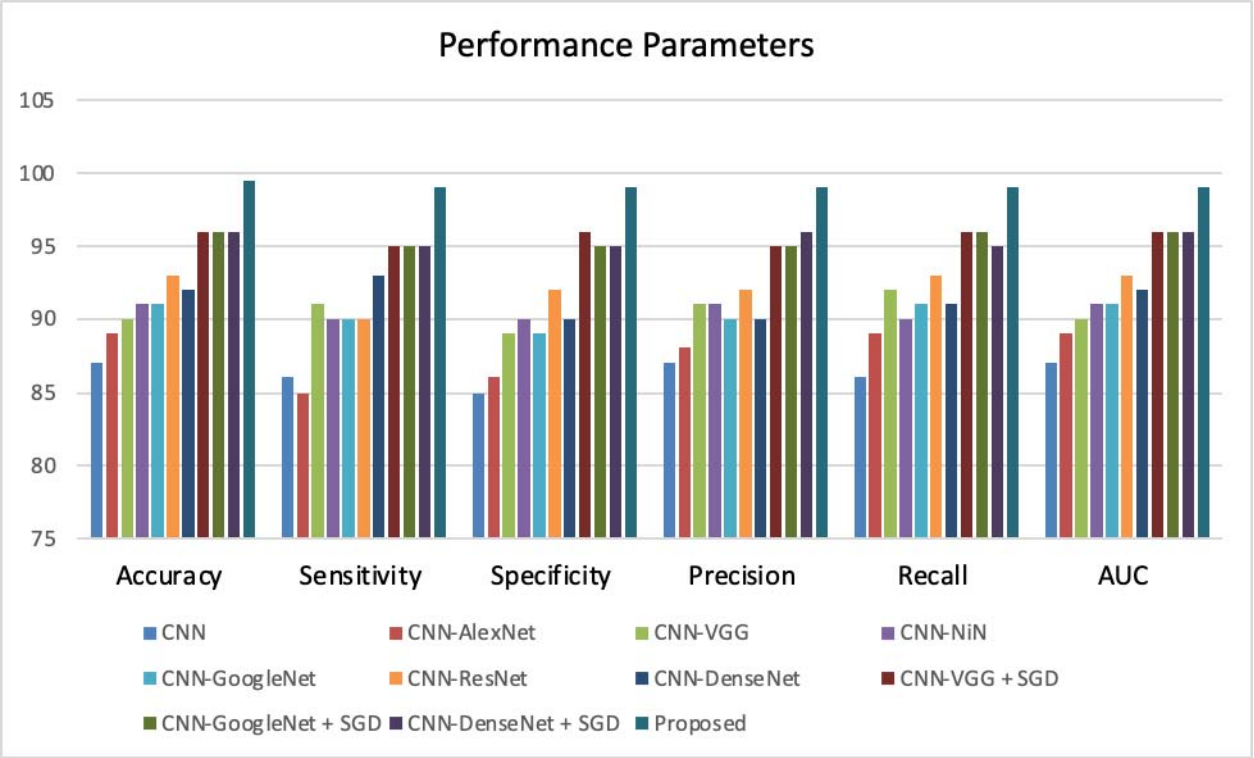


Figure 4.49. Performance parameters of different matrices with the proposed approach

19. The proposed work is compared with the work of (**Ghosh & Chatterjee, 2024, preprint)** evaluated ten ML algorithms (including SVM, XGBoost, CNN, RNN, Random Forest, and Logistic Regression) on UCI breast cancer data. SVM emerged best, achieving 98.25% accuracy, outperforming even CNN and XGBoost, as shown in the table (4-48) and figure (4-50).

Table 4.48.Performance parameters for different matrices

ML Algorithm	Accuracy (%)
XGBoost	0.9561
N N	0.9736
CNN	0.9473
RNN	0.9470
AdaBoost	0.9736
Adaptive Decision Learner	0.9385
LSTM Networks	0.9473
GRU	0.9385
R F	0.9649
SVM	0.9824
LR	0.9736
Proposed	**0.9947**

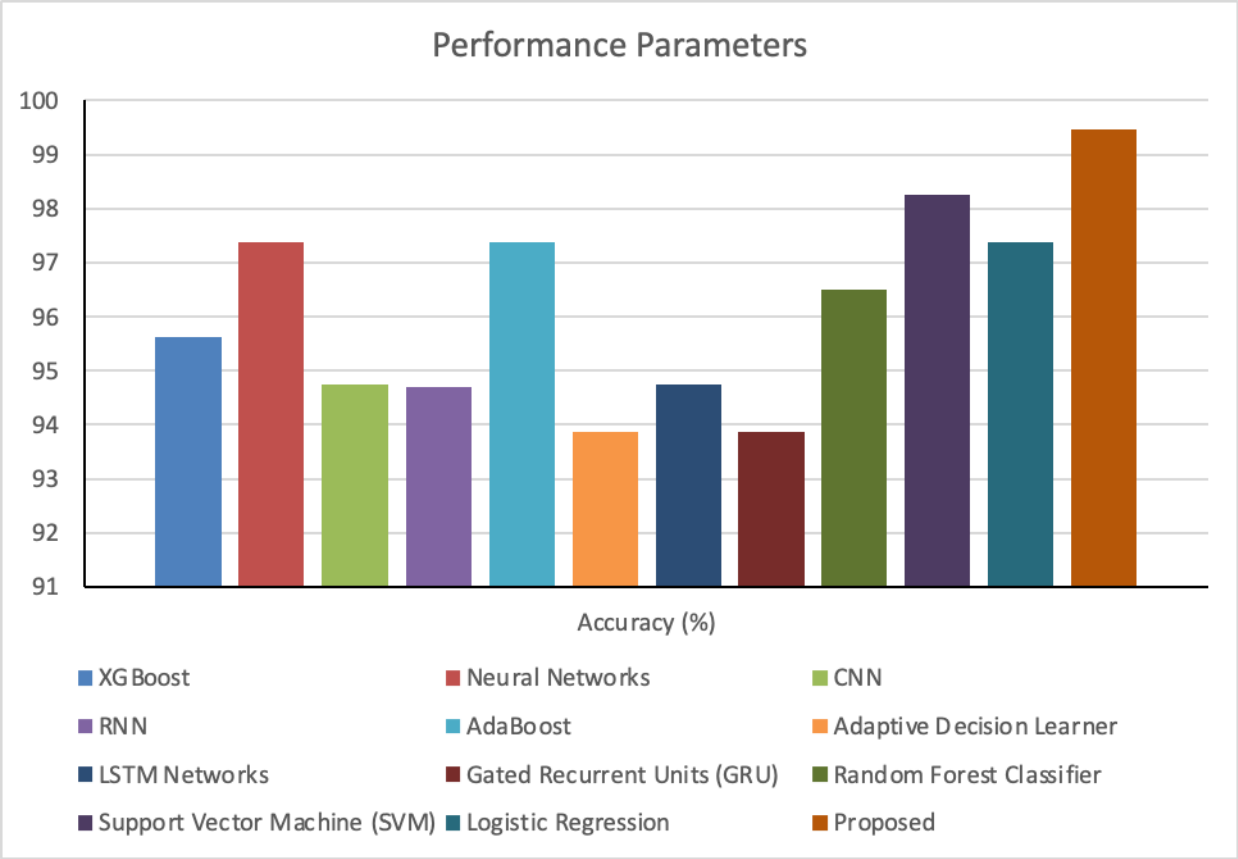


Figure 4.50. Accuracy at different matrices with the proposed approach

20. The proposed work is compared with the work of (Badar Almarri et al. 2024). Using machine learning techniques, including KNN, D.T., R.F., SVR, and Gaussian Naive Bayes (GNB), this research aims to create a breast cancer detection model. This research concludes that the Grid Search Algorithm outperforms the others, achieving 92.6383% accuracy, as shown in table (4-49) and figure (4-51).

Table 4.49. Performance parameters for different matrices

Model Name	Score	Accuracy
RF	0.9931	0.9256
DT	0.1000	0.9096
SVC	0.9238	0.9147
LR	0.9160	0.9097
Proposed	**0.9900**	**0.9947**

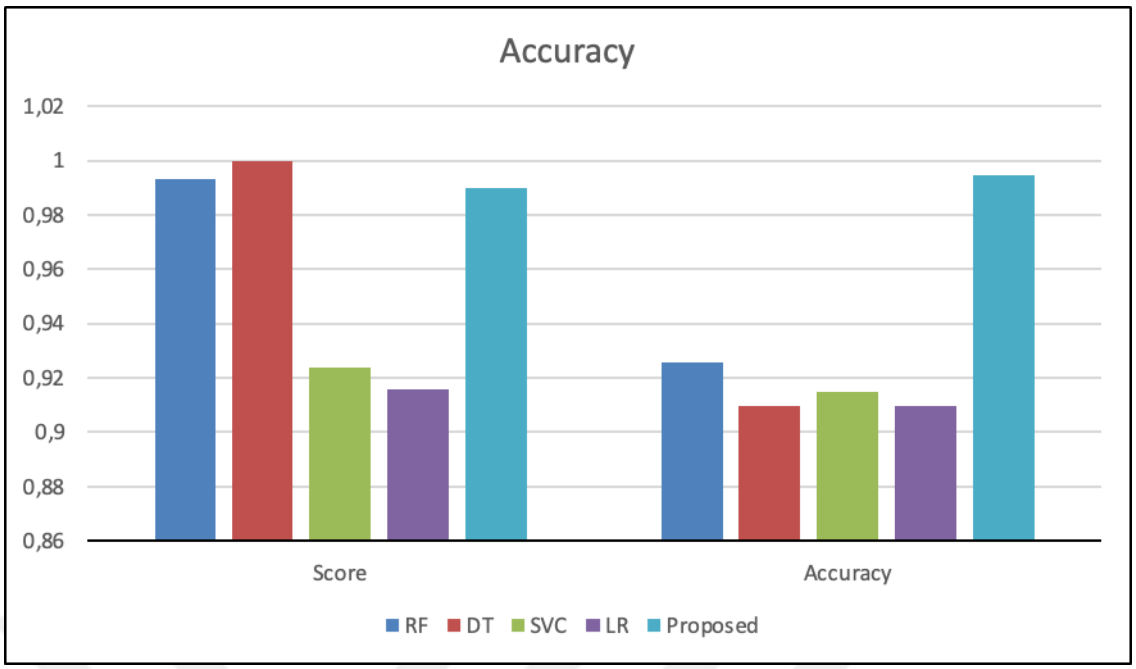


Figure 4.51. Accuracy of different matrices with the proposed approach

21. The proposed work is compared with the work of (Sharma et al., 2025) compares Random Forest, Neural Networks, Logistic Regression, and tuned models using Randomized Search CV across WBCD and Coimbra datasets. Random Forest demonstrated superior accuracy and generalizability, as shown in the table (4-50) and figure (4-52).

Table 4.50. Performance parameters for different matrices

Model	Accuracy	F1 Score	Precision	Recall
LR (from scratch)	0.9035	0.8842	0.8571	0.9130
LR (sklearn)	0.9561	0.9462	0.9362	0.9565
RF (sklearn)	0.9386	0.9213	0.9535	0.8913
Neural Network	0.9561	0.9213	0.9556	0.9348
Proposed	**0.9947**	**0.9900**	**0.9900**	**0.9900**

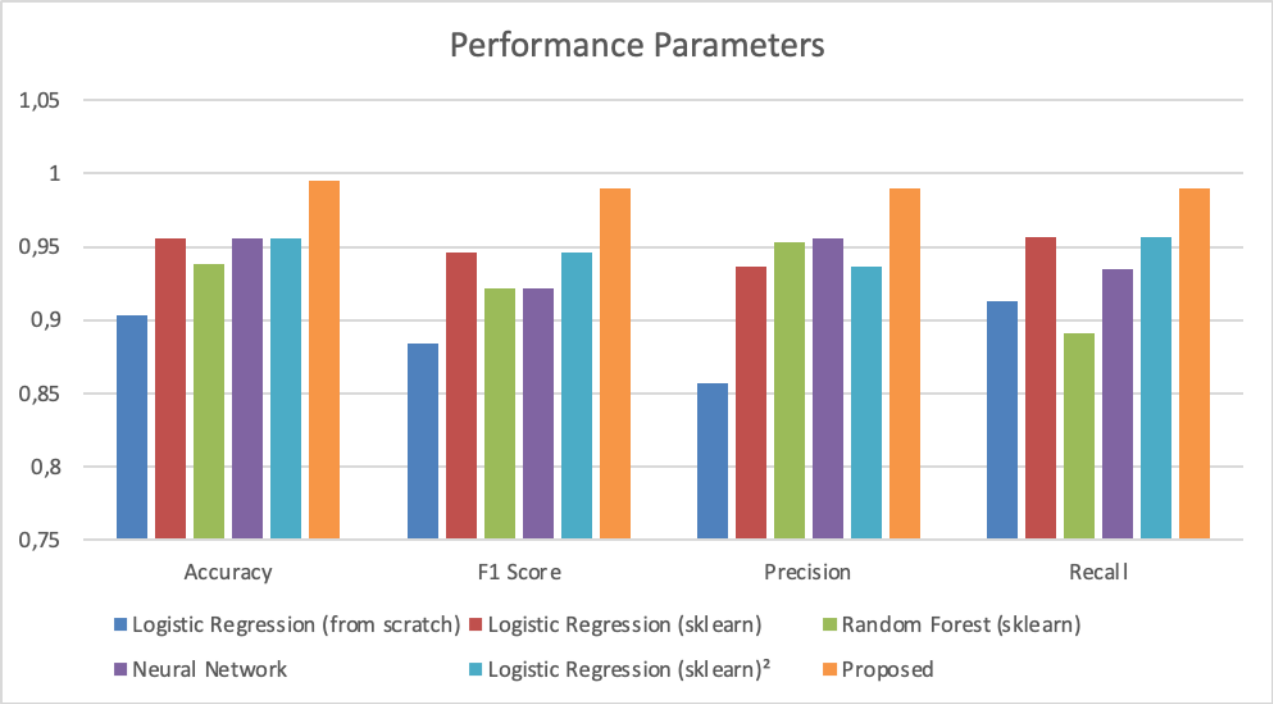


Figure 4.52. Accuracy of different matrices with the proposed approach

22. The proposed work is compared with the work of (M. Güler et al., 2025), DenseNet201, a deep convolutional neural network, was employed for breast cancer classification. The model achieved an accuracy of 89.4%, precision of 88.2%, recall of 84.1%, F1-score of 86.1%, and an AUC of 95.8%, underscoring the efficacy of deep learning in complex medical image analysis, as shown in the table (4-51) and figure (4-53).

Table 4.51. Performance parameters for different matrices

Model	Accuracy	Precision	Recall	F1 Score	AUC
Vanilla	0.8540	0.7880	0.8570	0.8210	0.9370
ResNet50	0.6090	0.0000	0.0000	0.0000	0.6020
ResNet152	0.6090	0.0000	0.0000	0.0000	0.7520
VGG16	0.7940	0.7900	0.6450	0.7100	0.8530
DenseNet152	0.8180	0.7850	0.7370	0.7600	0.8820
MobileNetv2	0.7730	0.6930	0.7550	0.7220	0.8560
EfficientB1	0.6090	0.0000	0.0000	0.0000	0.4720
NasNet	0.7050	0.6480	0.5370	0.5870	0.7620
DenseNet201	0.8940	0.8820	0.8410	0.8610	0.9580
Ensemble	0.8220	0.7280	0.8700	0.7930	0.9170
Tuned Model	0.8200	0.7620	0.7850	0.7730	0.8850
proposed	**0.9947**	**0.990**	**0.990**	**0.990**	**0.990**

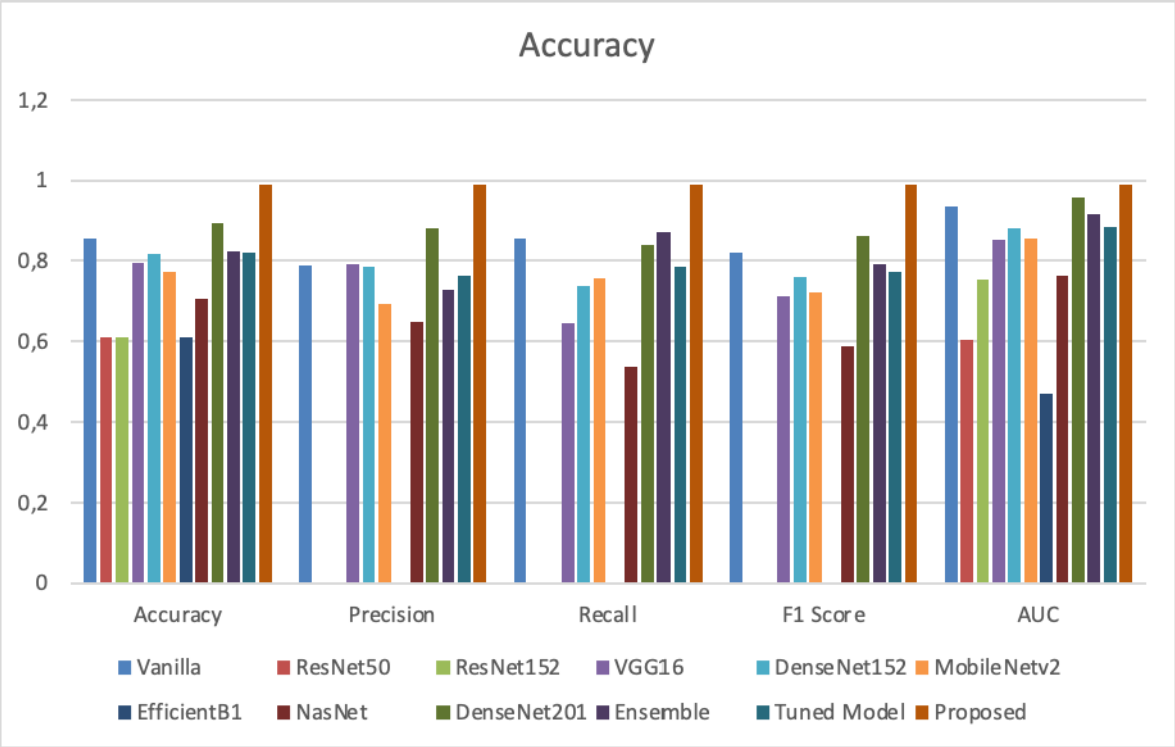


Figure 4.53. Accuracy of different matrices with the proposed approach

23. The proposed work is compared with the work of (H. Rafiepoor et al., 2025) compared various ML models, including Random Forest, Support Vector Machine, and Logistic Regression, with the traditional Gail risk assessment model for breast cancer risk prediction. The ML models outperformed the Gail model in recall and F1-score, highlighting the potential of AI-based approaches in early risk prediction, as shown in the table (4-52) and figure (4-54).

Table 4.52.Performance parameters for different matrices

Approach	Train Accuracy	Validation Accuracy	Sensitivity	Precision
Decision Tree	0.7008	0.5746	0.4786	0.5743
KNN	0.6598	0.6369	0.4914	0.6725
SVM	0.6455	0.6348	0.3760	0.7457
Random Forest	0.6954	0.6327	0.4829	0.6686
Bagging - DT	0.7088	0.62863	0.5512	0.6354
Bagging - SVM	0.642	0.62863	0.5512	0.6354
Gradient Boosting	0.6669	0.64522	0.5170	0.6759
AdaBoost	0.6473	0.64730	0.5384	0.6702
XGBoost	0.7061	0.61618	0.4700	0.6432
Rostami et al. [Ref]	0.6300	—	—	—
proposed	**0.9947**	**0.9900**	**0.9900**	**0.9900**

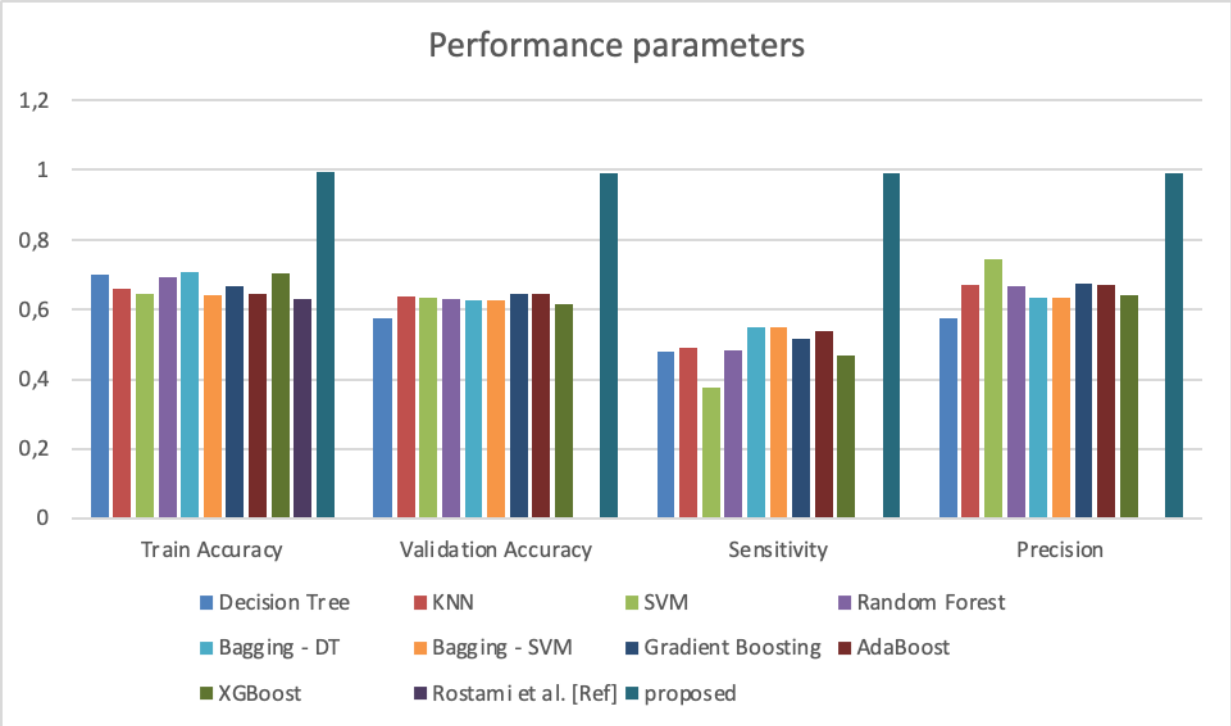


Figure 4.54. Performance parameters of different matrices with the proposed approach

24. The proposed work is compared with the work of (T. Arravalli et al., 2025), which utilized Random Forest classifiers combined with explainable AI techniques for breast cancer detection. The model achieved an F1-score of 84%, and integrating explainable AI methods provided transparency and interpretability for the model's predictions, as shown in table (4-53) and figure (4-55).

Table 4.53. Performance parameters for different matrices

Algorithm	Accuracy	Precision	Recall	F1-score	AUC
R F	0.8900	0.9300	0.9000	0.9300	0.9600
L R	0.8900	0.9000	0.8400	0.8700	0.9400
D T	0.8000	0.8300	0.7900	0.7100	0.9200
KNN	0.7800	0.7900	0.8300	0.7600	0.9300
Adaboost	0.8600	0.8300	0.7900	0.8100	0.9400
Catboost	0.8900	0.8100	0.8300	0.8400	0.9300
LightGBM	0.8300	0.8100	0.7900	0.8400	0.8000
XGBoost	0.8400	0.9000	0.8300	0.9300	0.9000
STACK	0.9300	0.9300	0.9000	0.9300	0.9200
proposed	**0.9947**	**0.9900**	**0.9900**	**0.9900**	**0.9900**

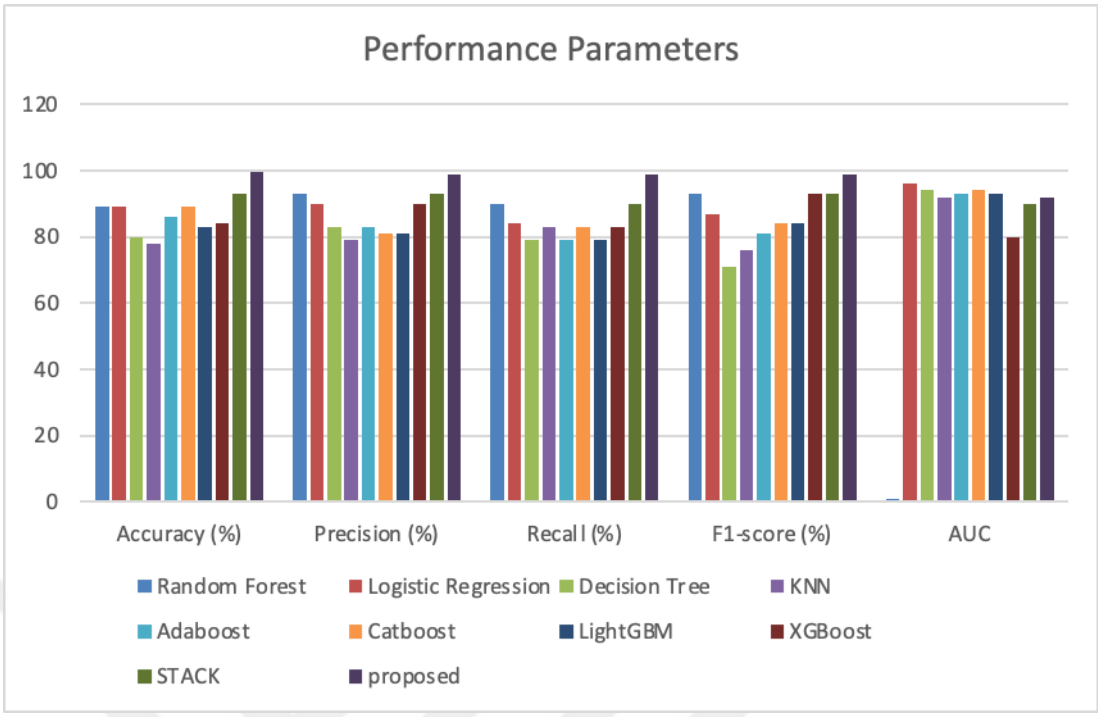


Figure 4.55. Performance parameters of different matrices with the proposed approach

25. The proposed work is compared with the work of (Umesh Kumar Lilhore et al. 2025 The study named "Hybrid convolutional neural network and bi-LSTM model with EfficientNet-B0 for high-accuracy breast cancer detection and classification" used CNN + Bi-LSTM + EfficientNet-B0 and achieved an accuracy score of 99.2%, as shown in the table (4-54) and figure (4-56).

Table 4.54. Performance parameters for different matrices

Model Variation	Accuracy	Precision	Recall	F1-Score
Full Model (CNN + EfficientNet-B0 + Bi-LSTM)	0.9920	0.9870	0.9950	0.9910
CNN + EfficientNet-B0 (Without Bi-LSTM)	0.9750	0.9680	0.9800	0.9740
CNN + Bi-LSTM (Without EfficientNet-B0)	0.9580	0.9450	0.9620	0.9530
CNN Only (Without Bi-LSTM or EfficientNet-B0)	0.9160	0.8940	0.9210	0.9070
Full Model (CNN + EfficientNet-B0 + Bi-LSTM)	0.9890	0.9800	0.9910	0.9860
Full Model (CNN + EfficientNet-B0 + Bi-LSTM)	0.9720	0.9600	0.9750	0.9670
Proposed	**0.9947**	**0.9900**	**0.9900**	**0.9900**

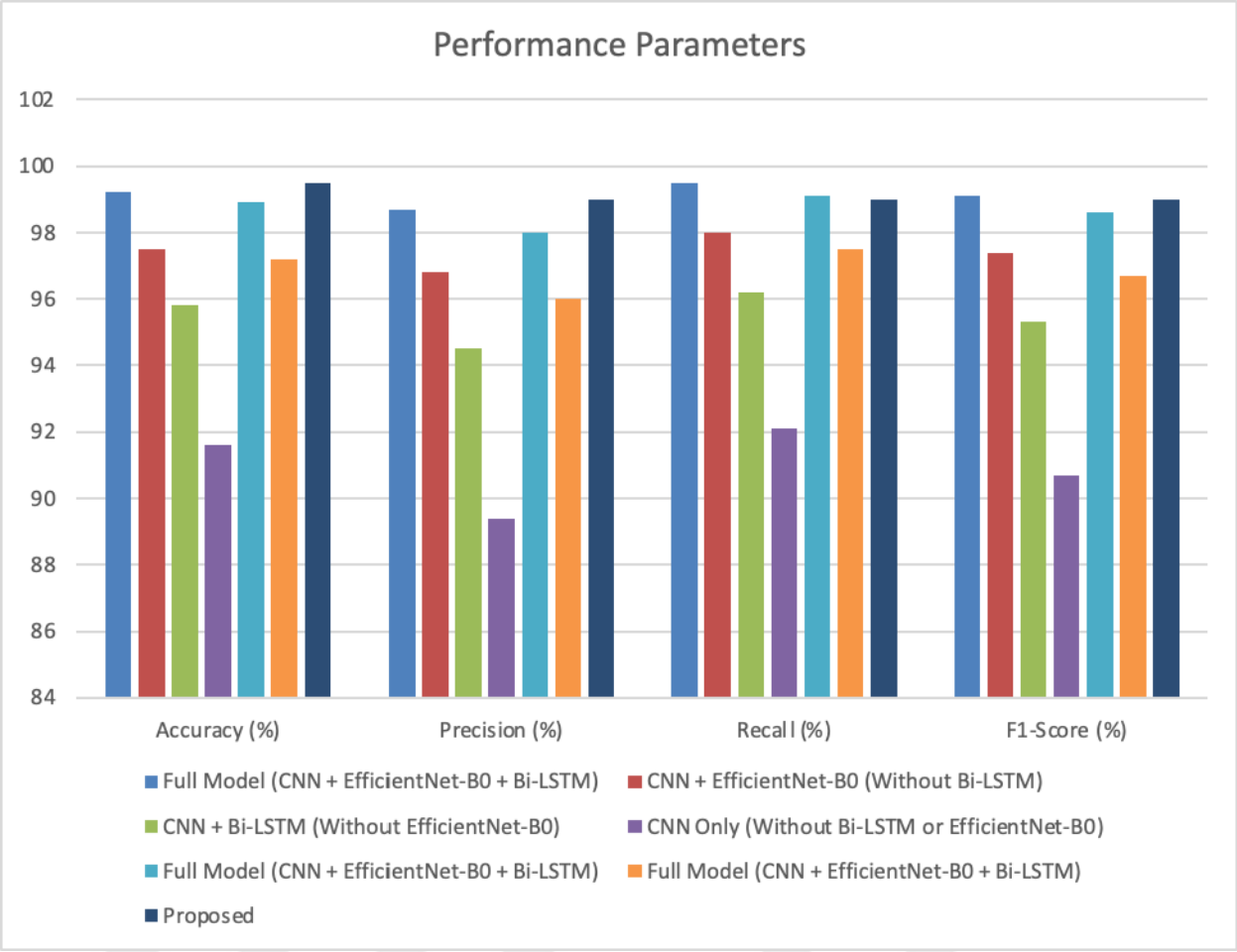


Figure 4.56. Performance parameters of different matrices with the proposed approach

It is essential to emphasize that the comparison between the findings of the present study and those reported in previous research does not require the hyperparameters or experimental settings to be identical across studies. Each investigation is conducted within its own methodological framework, characterized by differences in dataset composition, preprocessing strategies, model architectures, and tuning procedures. Consequently, variations in hyperparameters—such as dropout rate, number of hidden units, batch size, data-splitting ratios, and optimization algorithms—are both expected and inherent to the nature of machine learning and deep learning research. Therefore, the purpose of the comparative analysis presented herein is not to replicate the exact technical configurations employed in other studies, but rather to situate the outcomes of the current work within the broader scientific context and to elucidate their relative position in the existing body of literature. Moreover, variations in predictive performance across studies are considered normal and acceptable, given the diversity of experimental conditions and model-specific settings. Accordingly, the comparisons provided focus on substantive analytical insights while acknowledging that differences in hyperparameter configurations may contribute to the observed discrepancies in accuracy.

5. CONCLUSION AND FUTURE WORK

This research aims to develop an intelligent framework for improving breast cancer diagnosis based on deep machine learning techniques by designing a hybrid model that combines a Deep Multilayer Perceptron (Deep MLP), a Feature-Fused Autoencoder Neural Network, and a Weight-Tuned Cross-Validated Decision Tree.

The model was applied to the Wisconsin Diagnostic Breast Cancer (WDBC) dataset to achieve high diagnostic performance while optimizing the fine-tuning of model hyperparameters such as weights, data splitting ratios, batch size, number of hidden units, and dropout rate.

In addition, this study demonstrated the effectiveness of integrating clinical and genomic features from the METABRIC dataset to develop reliable and generalizable predictive models for long-term breast cancer outcomes. The results highlight the capability of the proposed hybrid algorithms and deep learning techniques to extract meaningful patterns from high-dimensional clinical–genomic data and to enhance prognostic accuracy. Overall, the findings confirm the value of combining multi-modal biomedical datasets and advanced hybrid deep learning architectures to support clinical decision-making and advance precision oncology.

The results showed that the proposed model achieved superior performance compared to traditional models and previous studies. The most important results can be summarized as follows:

- **Effect of Weights:**
  An improvement in classification accuracy was observed with increasing weight values up to a certain point, after which the accuracy began to fluctuate.
  The highest accuracy (0.958) was achieved at a weight of 9 with a 25% test set, which illustrates the importance of balancing overfitting and underfitting.

- **Effect of Train-Test Split Ratios:**
  Even with a reduced training data ratio, the model maintained strong performance, with an accuracy of 99.47% at a test ratio of 10% and an Area Under the Curve (AUC) of 0.996, demonstrating the model's high generalization ability.

- **Effect of Batch Size:**
  Performance metrics improved with increasing batch size up to a value of 128, where the best accuracy of 99.47% was recorded. Performance decreased slightly at larger batch sizes due to the increased likelihood of overfitting.

- **Effect of the Number of Hidden Units:**
  Accuracy gradually increased with the number of hidden units up to 256–512, where the model achieved an accuracy of 99.12%, reflecting a balance between computational complexity and the model's generalization ability.

- **Effect of Dropout Rate:**
  The best performance was achieved at a dropout rate of 0.45, where the model achieved an accuracy of 99.12% and an AUC value of 0.99. Performance decreased at higher dropout rates due to a reduction in the model's learning capacity.

In comparison with previous studies (from 2020 to 2025), the proposed model clearly outperformed all previous traditional and deep learning methods.

While the accuracies of other models ranged between 86% and 99.2%, the proposed model achieved an overall accuracy of 99.47% with precision, recall, and F1-score all reaching 99%, surpassing all previous models.

These results confirm that integrating deep learning techniques, autoencoders, and decision-making algorithms with optimal parameter tuning can produce a robust and effective diagnostic model capable of supporting physicians in early diagnosis and supporting e-Health systems.

Future Work

Despite the high performance achieved by the proposed model, several areas for future research could further enhance its efficiency and expand its scope of application, including:

1. **Expanding the Database and Diseases Covered:**
   Studying the application of the model to other diseases, such as thyroid cancer and liver diseases, and using multi-class datasets instead of binary ones.

2. **Practical Clinical Validation:**
   Collaborating with healthcare institutions and national bodies such as the National e-Health Authority (NEHTA) to apply the model to real patient data and test its accuracy in a real-world clinical setting.

3. **Explainable Artificial Intelligence (Explainable AI):**

Adding explanation and interpretation techniques to the model's decisions to make its outputs more transparent and understandable to physicians and specialists.

4. **Improving Computational Efficiency:**

Measuring performance according to additional criteria such as execution speed, computational cost, and resource consumption to develop a fast and responsive practical system that can be integrated into real-time medical applications.

5. **Transfer and Meta Learning:**

Using transfer learning and meta-learning techniques to improve the model's performance on small or new datasets without the need for complete retraining.

6. **Integrating the System into the e-Health Infrastructure:**

Connecting the proposed system with electronic health records (EHRs) to enable continuous monitoring, case analysis, and the provision of intelligent recommendations to physicians.

7. **Developing Deeper Hybrid Models:**

Exploring the integration of deep learning networks (such as CNN + Transformer) to create more accurate and flexible systems for analyzing both medical image and text data.

In general conclusion, this research demonstrated that the integration of deep neural networks, autoencoders, and weight-tuned decision-making algorithms significantly improves the accuracy of breast cancer diagnosis.

It also showed that fine-tuning hyperparameters (such as weights, batch size, hidden units, and dropout rate) plays a crucial role in achieving a robust and reliable generalizable model.

REFERENCES

J. C. Lashof, I. C. Henderson, and S. J. Nass, "Mammography and beyond: developing technologies for the early detection of breast cancer," 2001.

F. Rehman, "EARLY DETECTION OF BREAST CANCER: INNOVATIVE SCREENING METHODS," *Health and Medical Discoveries,* vol. 1, no. 02, pp. 50-72, 2024.

A. Pulumati, A. Pulumati, B. S. Dwarakanath, A. Verma, and R. V. Papineni, "Technological advancements in cancer diagnostics: Improvements and limitations," *Cancer Reports,* vol. 6, no. 2, p. e1764, 2023.

G. D. Schiff *et al.*, "Diagnosing diagnosis errors: lessons from a multi-institutional collaborative project," *Advances in patient safety: from research to implementation (volume 2: concepts and methodology),* 2005.

J. R. Ball, B. T. Miller, and E. P. Balogh, "Improving diagnosis in health care," 2015.

M. G. Hanna *et al.*, "Future of artificial intelligence (AI)-machine learning (ML) trends in pathology and medicine," *Modern Pathology,* p. 100705, 2025.

L. Wei *et al.*, "Artificial intelligence (AI) and machine learning (ML) in precision oncology: a review on enhancing discoverability through multiomics integration," *The British journal of radiology,* vol. 96, no. 1150, p. 20230211, 2023.

N. Kalra, P. Verma, and S. Verma, "Advancements in AI based healthcare techniques with FOCUS ON diagnostic techniques," *Computers in Biology and Medicine,* vol. 179, p. 108917, 2024.

M. Botlagunta *et al.*, "Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms," *Scientific Reports,* vol. 13, no. 1, p. 485, 2023.

A. M. Sharafaddini, K. K. Esfahani, and N. Mansouri, "Deep learning approaches to detect breast cancer: a comprehensive review," *Multimedia Tools and Applications,* vol. 84, no. 21, pp. 24079-24190, 2025.

S. J. S. Gardezi, A. Elazab, B. Lei, and T. Wang, "Breast cancer detection and diagnosis using mammographic data: Systematic review," *Journal of medical Internet research,* vol. 21, no. 7, p. e14464, 2019.

M. Ahsan, A. Khan, K. R. Khan, B. B. Sinha, and A. Sharma, "Advancements in medical diagnosis and treatment through machine learning: A review," *Expert systems,* vol. 41, no. 3, p. e13499, 2024.

S. Saturi, "Review on machine learning techniques for medical data classification and disease diagnosis," *Regenerative Engineering and Translational Medicine,* vol. 9, no. 2, pp. 141-164, 2023.

H. Abdel-Jaber, D. Devassy, A. Al Salam, L. Hidaytallah, and M. El-Amir, "A review of deep learning algorithms and their applications in healthcare," *Algorithms,* vol. 15, no. 2, p. 71, 2022.

O. Panahi, "Deep Learning in Diagnostics," *Journal of Medical Discoveries,* vol. 2, no. 1, pp. 1-6, 2025.

T. M. Pope, "Patient decision aids improve patient safety and reduce medical liability risk," *Me. L. Rev.,* vol. 74, p. 73, 2022.

B. Zhang, L. Yan, M. A. Moses, and P. C. Tsang, "Bovine membrane-type 1 matrix metalloproteinase: molecular cloning and expression in the corpus luteum," *Biology of reproduction,* vol. 67, no. 1, pp. 99-106, 2002.

S. Sriharikrishnaa, P. S. Suresh, and S. Prasada K, "An introduction to fundamentals of cancer biology," in *Optical Polarimetric modalities for biomedical research*: Springer, 2023, pp. 307-330.

P. Bisoyi, "Malignant tumors–as cancer," in *Understanding Cancer*: Elsevier, 2022, pp. 21-36.

J. M. Rajput, "Cancer: A comprehensive review," *International Journal of Research in Pharmacy and Allied Science,* vol. 1, no. 3, pp. 42-49, 2023.

N. K. Roy, D. Bordoloi, J. Monisha, A. Anip, G. Padmavathi, and A. B. Kunnumakkara, "Cancer—an overview and molecular alterations in cancer," *Fusion genes and Cancer,* pp. 1-15, 2017.

M. Akram, M. Iqbal, M. Daniyal, and A. U. Khan, "Awareness and current knowledge of breast cancer," *Biological research,* vol. 50, no. 1, p. 33, 2017.

M. Milosevic, D. Jankovic, A. Milenkovic, and D. Stojanov, "Early diagnosis and detection of breast cancer," *Technology and Health Care,* vol. 26, no. 4, pp. 729-759, 2018.

A. Mishra, A. Mandal, and S. Deo, "Clinical Approach to Symptomatic Breast and Triple Assessment," in *Imaging in Management of Breast Diseases: Volume 1, Overview of Modalities*: Springer, 2025, pp. 25-33.

E. Y. Trofimova, "An ultrasound of the lymph nodes will show whether there is inflammation in the body? Ultrasound of lymph nodes in the armpit and abdominal cavity."

J. Dixon and R. Mansel, "ABC of breast diseases: symptoms assessment and guidelines for referral," *Bmj,* vol. 309, no. 6956, pp. 722-726, 1994.

J. Zgajnar, "Clinical presentation, diagnosis and staging of breast cancer," in *Breast Cancer Management for Surgeons: A European Multidisciplinary Textbook*: Springer, 2017, pp. 159-176.

S. P. Na and D. Houserkovaa, "The role of various modalities in breast imaging," *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub,* vol. 151, no. 2, pp. 209-218, 2007.

E. P. Weledji and J. Tambe, "Breast cancer detection and screening," *Med Clin Rev,* vol. 4, no. 2, p. 8, 2018.

P. Priyadarshini, S. Sarathi, and V. Hemavathy, "Breast Cancer Screening," *Cardiometry,* no. 24, pp. 1000-1005, 2022.

A. M. Alalomari, "Pathologic Findings from Breast Screening: Outcomes in Females Who Had a Biopsy After Mammography Screening," Alfaisal University (Saudi Arabia), 2024.

I. Andersson, "Invasive breast cancer," in *Radiologic-Pathologic Correlations from Head to Toe: Understanding the Manifestations of Disease*: Springer, 2005, pp. 757-766.

E. B. Effiong, B. C. Nweke, A. H. Okereke, and A. N. O. Pius, "BREAST CANCER–REVIEW," *Journal of Global Biosciences Vol,* vol. 11, no. 3, pp. 9248-9257, 2022.

G. Cserni, "Histological type and typing of breast carcinomas and the WHO classification changes over time," *Pathologica,* vol. 112, no. 1, p. 25, 2020.

S. Zhao, W.-J. Zuo, Z.-M. Shao, and Y.-Z. Jiang, "Molecular subtypes and precision treatment of triple-negative breast cancer," *Annals of translational medicine,* vol. 8, no. 7, p. 499, 2020.

G. K. Reeves, K. Pirie, J. Green, D. Bull, and V. Beral, "Reproductive factors and specific histological types of breast cancer: prospective study and meta-analysis," *British journal of cancer,* vol. 100, no. 3, pp. 538-544, 2009.

S. Darb-Esfahani *et al.*, "Identification of biology-based breast cancer types with distinct predictive and prognostic features: role of steroid hormone and HER2 receptor expression in patients treated with neoadjuvant anthracycline/taxane-based chemotherapy," *Breast cancer research,* vol. 11, no. 5, p. R69, 2009.

X. Dai, A. Chen, and Z. Bai, "Integrative investigation on breast cancer in ER, PR and HER2-defined subgroups using mRNA and miRNA expression profiling," *Scientific reports,* vol. 4, no. 1, p. 6566, 2014.

P. A. Newcomb and P. P. Carbone, "Cancer treatment and age: patient perspectives," *JNCI: Journal of the National Cancer Institute,* vol. 85, no. 19, pp. 1580-1584, 1993.

B. O. Anderson, R. Masetti, and M. J. Silverstein, "Oncoplastic approaches to partial mastectomy: an overview of volume-displacement techniques," *The lancet oncology,* vol. 6, no. 3, pp. 145-157, 2005.

D. A. Mahvi, R. Liu, M. W. Grinstaff, Y. L. Colson, and C. P. Raut, "Local cancer recurrence: the realities, challenges, and opportunities for new therapies," *CA: a cancer journal for clinicians,* vol. 68, no. 6, pp. 488-505, 2018.

V. Hearnden *et al.*, "New developments and opportunities in oral mucosal drug delivery for local and systemic disease," *Advanced drug delivery reviews,* vol. 64, no. 1, pp. 16-28, 2012.

M. Benderitter *et al.*, "Stem cell therapies for the treatment of radiation-induced normal tissue side effects," *Antioxidants & redox signaling,* vol. 21, no. 2, pp. 338-355, 2014.

R. Kaur, A. Bhardwaj, and S. Gupta, "Cancer treatment therapies: traditional to modern approaches to combat cancers," *Molecular biology reports,* vol. 50, no. 11, pp. 9663-9676, 2023.

C. Russell, "Contributing Factors to Mammography Screening among African American and Hispanic Women: Quantitative Correlation Study," Walden University, 2022.

C. E. DeSantis *et al.*, "Breast cancer statistics, 2019," *CA: a cancer journal for clinicians,* vol. 69, no. 6, pp. 438-451, 2019.

B. Stenkvist *et al.*, "Predicting breast cancer recurrence," *Cancer,* vol. 50, no. 12, pp. 2884-2893, 1982.

A. Mustafa and M. Rahimi Azghadi, "Automated machine learning for healthcare and clinical notes analysis," *Computers,* vol. 10, no. 2, p. 24, 2021.

W. W. Harless, "Cancer treatments transform residual cancer cell phenotype," *Cancer cell international,* vol. 11, no. 1, p. 1, 2011.

A. S. Abdullah, A. G. Ahmed, S. N. Mohammed, A. A. Qadir, N. M. Bapir, and G. M. Fatah, "Benign tumor publication in one year (2022): a cross-sectional study," *Barw Medical Journal,* 2023.

F. W. Foote and F. W. Stewart, "Comparative studies of cancerous versus noncancerous breasts," *Annals of surgery,* vol. 121, no. 1, pp. 6-53, 1945.

P. Bisoyi, "A brief tour guide to cancer disease," in *Understanding cancer*: Elsevier, 2022, pp. 1-20.

R. Rajendran, "Benign and malignant tumors of the oral cavity," *Shafer's Textbook of Oral Pathology,* vol. 6, pp. 169-73, 2009.

J. Wang *et al.*, "Progression from ductal carcinoma in situ to invasive breast cancer: molecular features and clinical significance," *Signal transduction and targeted therapy,* vol. 9, no. 1, p. 83, 2024.

A. Bang and S. Nagpure, "Determination of Risk Factors for DCIS Ductal in-situ Carcinoma of Mammary Gland," *Journal of Pharmaceutical Research International,* vol. 33, no. 60B, pp. 757-763, 2021.

A. E. Gallas *et al.*, "An Overview of Invasive Ductal Carcinoma (IDC) in Women's Breast Cancer," *Current Molecular Medicine,* 2025.

E. R. Shamir, H. Hwang, and Y.-Y. Chen, "Invasive lobular carcinoma," in *A Comprehensive Guide to Core Needle Biopsies of the Breast*: Springer, 2022, pp. 655-690.

M. Christgen *et al.*, "Lobular breast cancer: histomorphology and different concepts of a special spectrum of tumors," *Cancers,* vol. 13, no. 15, p. 3695, 2021.

A. E. McCart Reed, L. Kalinowski, P. T. Simpson, and S. R. Lakhani, "Invasive lobular carcinoma of the breast: the increasing importance of this special subtype," *Breast Cancer Research,* vol. 23, no. 1, p. 6, 2021.

E. B. Ikhuoria and C. Bach, "Introduction to breast carcinogenesis–symptoms, risks factors, treatment and management," *European Journal of Engineering and Technology Research,* vol. 3, no. 7, pp. 58-66, 2018.

T. Abd Alkareem, S. Hassan, and S. Abdalhadi, "Breast cancer: symptoms, causes, and treatment by metal complexes: a review," *Advanced Journal of Chemistry-Section B: Natural Products and Medical Chemistry,* vol. 5, no. 4, pp. 306-319, 2023.

W. Lu, L. Jansen, W. Post, J. Bonnema, J. Van de Velde, and G. De Bock, "Impact on survival of early detection of isolated breast recurrences after the primary treatment for breast cancer: a meta-analysis," *Breast cancer research and treatment,* vol. 114, no. 3, pp. 403-412, 2009.

N. Houssami, S. Ciatto, F. Martinelli, R. Bonardi, and S. Duffy, "Early detection of second breast cancers improves prognosis in breast cancer survivors," *Annals of oncology,* vol. 20, no. 9, pp. 1505-1510, 2009.

N. Kumar, G. Sharma, and L. Bhargava, "The machine learning based optimized prediction method for breast cancer detection," in *2020 4th International conference on electronics, communication and aerospace technology (ICECA)*, 2020: IEEE, pp. 1594-1598.

G. Battineni, N. Chintalapudi, and F. Amenta, "Performance analysis of different machine learning algorithms in breast cancer predictions," *EAI Endorsed Transactions on Pervasive Health & Technology,* vol. 6, no. 23, 2020.

P. Karatza, K. Dalakleidi, M. Athanasiou, and K. S. Nikita, "Interpretability methods of machine learning algorithms with applications in breast cancer diagnosis," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021: IEEE, pp. 2310-2313.

J. Wu and C. Hicks, "Breast cancer type classification using machine learning," *Journal of personalized medicine,* vol. 11, no. 2, p. 61, 2021.

V. P. T. Vy, M. M.-S. Yao, N. Q. Khanh Le, and W. P. Chan, "Machine learning algorithm for distinguishing ductal carcinoma in situ from invasive breast cancer," *Cancers,* vol. 14, no. 10, p. 2437, 2022.

M. A. Naji, S. El Filali, K. Aarika, E. H. Benlahmar, R. Ait Abdelouhahid, and O. Debauche, "Machine learning algorithms for breast cancer prediction and diagnosis," *Procedia Computer Science,* vol. 191, pp. 487-492, 2021.

A. I. Aishwarja, N. J. Eva, S. Mushtary, Z. Tasnim, N. I. Khan, and M. N. Islam, "Exploring the machine learning algorithms to find the best features for predicting the breast cancer and its recurrence," in *International conference on intelligent computing & optimization*, 2020: Springer, pp. 546-558.

S. Sakib, N. Yasmin, A. K. Tanzeem, F. Shorna, K. Md. Hasib, and S. B. Alam, "Breast cancer detection and classification: A comparative analysis using machine learning algorithms," in *Proceedings of third international conference on communication, computing and electronics systems: ICCCES 2021*, 2022: Springer, pp. 703-717.

B. Thakur, N. Kumar, and G. Gupta, "Machine learning techniques with ANOVA for the prediction of breast cancer," *International Journal of Advanced Technology and Engineering Exploration,* vol. 9, no. 87, pp. 232-245, 2022.

M. Monirujjaman Khan *et al.*, "Machine learning based comparative analysis for breast cancer prediction," *Journal of Healthcare Engineering,* vol. 2022, 2022.

V. Nemade and V. Fegade, "Machine Learning Techniques for Breast Cancer Prediction," *Procedia Computer Science,* vol. 218, pp. 1314-1320, 2023.

K. M. M. Uddin, N. Biswas, S. T. Rikta, and S. K. Dey, "Machine learning-based diagnosis of breast cancer utilizing feature optimization technique," *Computer Methods and Programs in Biomedicine Update,* vol. 3, p. 100098, 2023.

P. Dinesh and P. Kalyanasundaram, "Medical image prediction for the diagnosis of breast cancer and comparing machine learning algorithms: SVM, logistic regression, random forest and decision tree to measure accuracy of prediction," in *AIP Conference Proceedings*, 2023, vol. 2821, no. 1: AIP Publishing LLC, p. 050001.

A. Fatima *et al.*, "Analyzing breast cancer detection using machine learning & deep learning techniques," *Journal of Computing & Biomedical Informatics,* vol. 7, no. 02, 2024.

R. Zeng *et al.*, "FastLeakyResNet-CIR: A novel deep learning framework for breast cancer detection and classification," *IEEE Access,* vol. 12, pp. 70825-70832, 2024.

P. S. C. Murty *et al.*, "Integrative hybrid deep learning for enhanced breast cancer diagnosis: leveraging the Wisconsin Breast Cancer Database and the CBIS-DDSM dataset," *Scientific Reports,* vol. 14, no. 1, p. 26287, 2024.

P. Ghosh and D. Chatterjee, "Comparative analysis of machine learning algorithms for breast cancer classification: SVM outperforms XGBoost, CNN, RNN, and others," *bioRxiv,* p. 2024.04. 22.590658, 2024.

B. Almarri, G. Gupta, R. Kumar, V. Vandana, F. Asiri, and S. B. Khan, "The BCPM method: decoding breast cancer with machine learning," *BMC medical imaging,* vol. 24, no. 1, p. 248, 2024.

T. Sharma, M. Mishra, S. Sharma, V. P. Chaubey, and K. Krishan, "Enhancing Breast Cancer Diagnosis and Prognosis Through Machine Learning and Deep Learning: A Comparative Analysis of Predictive Models," in *International Conference on Advanced Informatics for Computing Research*, 2023: Springer, pp. 3-13.

M. Güler, G. Sart, Ö. Algorabi, A. N. Adıguzel Tuylu, and Y. S. Türkan, "Breast Cancer Classification with Various Optimized Deep Learning Methods," *Diagnostics,* vol. 15, no. 14, p. 1751, 2025.

H. Rafiepoor *et al.*, "Comparison of Machine Learning Models for Classification of Breast Cancer Risk Based on Clinical Data," *Cancer Reports,* vol. 8, no. 4, p. e70175, 2025.

T. Arravalli *et al.*, "Detection of breast cancer using machine learning and explainable artificial intelligence," *Scientific Reports,* vol. 15, no. 1, p. 26931, 2025.

U. K. Lilhore *et al.*, "Hybrid convolutional neural network and bi-LSTM model with EfficientNet-B0 for high-accuracy breast cancer detection and classification," *Scientific Reports,* vol. 15, no. 1, p. 12082, 2025.

T. A. Shaikh and R. Ali, "Applying machine learning algorithms for early diagnosis and prediction of breast cancer risk," in *Proceedings of 2nd International Conference on Communication, Computing and Networking: ICCCN 2018, NITTTR Chandigarh, India*, 2019: Springer, pp. 589-598.

G. Chugh, S. Kumar, and N. Singh, "Survey on machine learning and deep learning applications in breast cancer diagnosis," *Cognitive Computation,* pp. 1-20, 2021.

T. M. Shadman, F. S. Akash, and M. Ahmed, "Machine learning as an indicator for breast cancer prediction," BRAC University, 2018.

S. K. Karthikeyan *et al.*, "MammOnc-DB, an integrative breast cancer data analysis platform for target discovery," *Npj Breast Cancer,* vol. 11, no. 1, p. 35, 2025.

A. a. El-Nabawy, N. A. Belal, and N. El-Bendary, "A cascade deep forest model for breast cancer subtype classification using multi-omics data," *Mathematics,* vol. 9, no. 13, p. 1574, 2021.

S. E. Clare and P. L. Shaw, "“Big Data” for breast cancer: where to look and what you will find," *NPJ breast cancer,* vol. 2, no. 1, pp. 1-5, 2016.

R. G. Martinez and D.-M. van Dongen, "Deep learning algorithms for the early detection of breast cancer: A comparative study with traditional machine learning," *Informatics in Medicine Unlocked,* vol. 41, p. 101317, 2023.

D. Lakshmi, S. R. Gurrela, and M. Kuncharam, "A comparative study on breast cancer tissues using conventional and modern machine learning models," in *Smart Computing Techniques and Applications: Proceedings of the Fourth International Conference on Smart Computing and Informatics, Volume 1*, 2021: Springer, pp. 693-699.

S. Sahran *et al.*, "Machine learning methods for breast cancer diagnostic," in *Breast Cancer and Surgery*: IntechOpen, 2018.

Y. e. Gao, J. Lin, Y. Zhou, and R. Lin, "The application of traditional machine learning and deep learning techniques in mammography: a review," *Frontiers in Oncology,* vol. 13, p. 1213045, 2023.

J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data,* vol. 6, no. 1, pp. 1-54, 2019.

I. Goodfellow, "Deep learning," ed: MIT press, 2016.

T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE transactions on pattern analysis and machine intelligence,* vol. 41, no. 2, pp. 423-443, 2018.

M. Aswathy and M. Jagannath, "Performance analysis of segmentation algorithms for the detection of breast cancer," *Procedia Computer Science,* vol. 167, pp. 666-676, 2020.

K. G. Sheela and S. N. Deepa, "Review on methods to fix number of hidden neurons in neural networks," *Mathematical problems in engineering,* vol. 2013, 2013.

G. Battineni, N. Chintalapudi, and F. Amenta, "Performance analysis of different machine learning algorithms in breast cancer predictions," *EAI Endorsed Transactions on Pervasive Health and Technology,* vol. 6, no. 23, 2020.

A. I. Aishwarja, N. J. Eva, S. Mushtary, Z. Tasnim, N. I. Khan, and M. N. Islam, "Exploring the machine learning algorithms to find the best features for predicting the breast cancer and its recurrence," in *Intelligent Computing and Optimization: Proceedings of the 3rd International Conference on Intelligent Computing and Optimization 2020 (ICO 2020)*, 2021: Springer, pp. 546-558.

H. Prajapati, "Breast Cancer Classification using different Machine Learning Algorithms," CALIFORNIA STATE UNIVERSITY, NORTHRIDGE, 2023.

CURRICULUM VITAE

Nagham Rasheed Hasmeed ALSAEDI graduated from Al Mustansiriyah University, College of Education ,Department of Computer Scıence in 2007. She started her master's program in Department of Computer Engineering at Erciyes University ,Faculty of Engineering ,in 2016 and coputed it in 2018. She began her Ph.D.studies in Department of Engineering at Çukurova University faculty of engineering in 2019 and successfully computed them in 2025 .She worked at the Iraqi Ministry of Trasportation in the Computer Department ,and she is currently working at the General Commission for Taxes in the Computer Department.