Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.3, June 2026, pp. 2336~2348
e-ISSN: 2548-6861

Deep Learning and XGBoost for Pancreatic Cancer Survival Prediction: A Real-World Evaluation in a Resource-Constrained African Healthcare Setting

Zvinodashe Revesai [1]*, Maguraushe Kudakwashe [2] , Ndlovu Belinda [3]*****
[1]Reformed Church University, Masvingo, Zimbabwe
[2]School of Computing, University of South Africa, Pretoria, South Africa
[3]Informatics Department, National University of Science and Technology, Bulawayo, Zimbabwe
[1]revesaiz@rcu.ac.zw , [2]magark@unisa.ac.za , [3]belinda.ndlovu@nust.ac.zw

Article Info

Article history:

Received 2026-02-25
Revised 2026-05-06
Accepted 2026-05-18

Keyword:

Pancreatic Cancer Survival Prediction,
Artificial Neural Networks,
XGBoost,
Clinical Decision Support Systems,
Resource-Constrained Healthcare,
Machine Learning in Oncology,
African Healthcare Systems.

ABSTRACT

Pancreatic cancer remains one of the most lethal malignancies worldwide, with persistently low survival rates and a pressing need for reliable prognostic tools to support treatment planning in resource-constrained healthcare environments. This study presents a structured comparative evaluation of Artificial Neural Network (ANN) and XGBoost classifiers for predicting 12-month survival using real-world clinical data from 569 pancreatic cancer patients treated at a public hospital in Zimbabwe between 2018 and 2023. The Cross-Industry Standard Process for Data Mining (CRISP-DM) framework guided data understanding, preprocessing, model development, and evaluation. A comprehensive preprocessing pipeline incorporating missing value imputation, outlier management, encoding, feature selection, and normalisation was applied, with all transformations derived exclusively from the training set to prevent data leakage. Models were trained using an 80/20 stratified split with cross-validated hyperparameter optimisation and evaluated on a strictly held-out test set using accuracy, precision, recall, F1-score, ROC analysis, and McNemar's test. On the test dataset, the ANN model achieved 99% overall accuracy and 99% F1-score, outperforming XGBoost, which attained 90% accuracy and 90% F1-score. The performance difference was statistically significant ($p < 0.05$). Computational analysis demonstrated inference times below 3 milliseconds per sample, supporting feasibility for clinical deployment. While results indicate strong discriminative capacity within this single-centre dataset, external validation across multi-institutional cohorts is necessary to confirm generalisability. These findings suggest that supervised machine learning can provide clinically meaningful support for survival prediction in African tertiary healthcare settings. This study uniquely contributes a deployment-oriented, real-world evaluation of machine learning models within a resource-constrained African healthcare context, addressing a critical gap in the current oncology informatics literature.

I. INTRODUCTION

Pancreatic cancer represents one of the most challenging problems in medical predictive modelling due to its highly complex clinical presentation, aggressive progression patterns, and multifactorial aetiology. Globally, pancreatic cancer ranks as the seventh leading cause of cancer-related mortality, with an estimated 495,773 new cases and 466,003 deaths reported worldwide in 2020 [1], making it one of the few malignancies for which incidence and mortality rates remain nearly equivalent. The disease exhibits a five-year survival rate of approximately 9% [2][3], a figure that has improved only marginally over several decades despite substantial advances in oncological research and treatment

modalities. This persistently poor prognosis is attributable primarily to the absence of reliable early biomarkers, the aggressive biological behaviour of the tumour, and the high frequency of late-stage diagnosis at presentation, all of which collectively render accurate prognostic prediction critical for clinical decision support and resource optimisation in healthcare systems operating under constrained conditions.

From a computational perspective, predicting survival outcomes in pancreatic cancer patients constitutes a high-stakes binary classification task characterised by several inherent challenges. Datasets in this domain are characterised by truncated survival distributions and right-censored observations, which complicate standard classification frameworks [4]. The relevant feature space spans multiple data modalities, including demographic variables, laboratory measurements, and pathological characteristics, creating high-dimensional input vectors susceptible to the well-documented curse of dimensionality [5]. The underlying relationships between predictors and survival outcomes are inherently nonlinear and involve complex interactions that cannot be adequately captured by linear models or simple parametric assumptions [6]. These characteristics collectively necessitate sophisticated machine learning approaches capable of learning hierarchical representations and modelling complex decision boundaries across diverse clinical feature spaces.

Accurate survival prediction in pancreatic cancer serves multiple critical functions within clinical information systems [7]. Prognostic models enable automated risk stratification for patient cohorts, facilitate evidence-based treatment planning through decision support algorithms, and support personalised medicine initiatives through patient-specific outcome predictions. From a healthcare informatics perspective, robust survival prediction models can be integrated into electronic health record systems and clinical decision support platforms to enhance operational efficiency and clinical outcomes, a consideration of relevance in resource-constrained healthcare environments where the efficient allocation of limited clinical resources can have a direct and measurable impact on patient survival [7].

Traditional approaches to survival analysis have predominantly relied on statistical methods such as Cox proportional hazards regression [8] and Kaplan-Meier estimators [9]. While these classical techniques have provided valuable baseline models, they possess fundamental computational and theoretical limitations that restrict their applicability to complex real-world clinical datasets. Cox regression assumes proportional hazards and linear log-hazard functions, assumptions that may not hold in datasets characterised by nonlinear predictor-outcome relationships and time-varying covariate effects [10]. Kaplan-Meier methods provide nonparametric survival curve estimation but lack predictive capability for individual patients, limiting their utility as decision-support tools. These conventional models additionally require manual feature engineering, cannot automatically learn feature representations from raw data, and often fail to capture complex nonlinear interaction effects that characterise high-dimensional clinical data [10].

The exponential growth of electronic health records and advances in computational infrastructure have enabled the application of sophisticated machine learning algorithms to clinical prediction tasks [11]. Machine learning, particularly supervised learning paradigms, offers a data-driven framework for developing predictive models that can automatically discover patterns in complex datasets without requiring explicit programming of decision rules [12]. Modern machine learning algorithms can perform automatic feature learning, model arbitrary nonlinear decision boundaries, handle heterogeneous and high-dimensional data, and generalise effectively to unseen instances through regularisation and ensemble techniques, advantages that position them as natural successors to traditional statistical approaches in clinical prognostic modelling [6][12].

Within the supervised learning paradigm, Artificial Neural Networks (ANNs) employ multi-layer architectures with nonlinear activation functions to learn hierarchical feature representations through backpropagation-based optimisation [13]. Deep neural networks with multiple hidden layers can approximate arbitrary continuous functions, making them universal function approximators well-suited for modelling complex clinical relationships. Gradient boosting algorithms, particularly XGBoost, utilise ensemble learning principles by combining multiple weak learners through iterative boosting to construct robust predictive models, incorporating regularisation terms, native handling of missing data, and advanced second-order gradient optimisation, thereby achieving strong generalisation performance on structured tabular data [14]. Both algorithmic families have demonstrated strong performance across a range of cancer outcome prediction tasks [15], yet their comparative performance on pancreatic cancer survival data from resource-limited African healthcare settings remains insufficiently characterised in the existing literature.

Most published machine learning studies in oncology utilise datasets from high-income countries with advanced healthcare infrastructure, raising questions about the generalisability of these models to resource-constrained environments where data quality, feature availability, and patient populations may differ substantially [16]. Limited research has conducted a systematic comparative evaluation of multiple supervised learning architectures using identical real-world clinical datasets from such contexts, making it difficult to assess relative algorithmic performance under the specific conditions that characterise healthcare delivery in developing countries. This study addresses these research gaps by implementing and evaluating supervised machine learning models for pancreatic cancer survival prediction using real-world clinical data from a hospital in Zimbabwe, following the Cross-Industry Standard Process for Data Mining methodology as a systematic framework for model development and evaluation [17].

The main contributions of this work are as follows:

1. Artificial Neural Network and XGBoost classifiers are developed and trained for binary survival prediction using the CRISP-DM methodology applied to real-world clinical data from a resource-limited tertiary healthcare institution in Zimbabwe.
2. A systematic comparative evaluation of these models is conducted using multiple performance metrics on identical train-test splits, providing rigorous empirical evidence of relative algorithmic performance.
3. The feasibility of deploying machine learning-based clinical decision support systems in resource-constrained healthcare environments is assessed, with specific attention to computational efficiency and practical deployment requirements.
4. The optimal supervised learning approach is identified based on empirical performance evaluation, providing a practical recommendation for potential clinical integration within the target healthcare setting.

The remainder of this paper is organised as follows: Section 2 reviews related work on machine learning approaches to cancer survival prediction. Section 3 describes the dataset and proposed methodology, including preprocessing and model development. Section 4 presents experimental results. Section 5 discusses the findings and their implications. Section 6 provides concluding remarks and directions for future work.

II. LITERATURE REVIEW

A. Traditional Statistical Methods for Survival Analysis

Traditional survival analysis relies on statistical methods such as Cox proportional hazards regression and Kaplan-Meier estimators that have formed the backbone of clinical prognostic modelling for several decades [8][9]. Cox regression models the hazard function as a product of a baseline hazard and an exponential linear predictor, enabling the simultaneous evaluation of multiple covariates on survival outcomes while assuming proportional hazards over time [8]. While this framework has provided a principled statistical foundation for clinical prognosis, it imposes strict parametric assumptions, including linearity of log-hazards and independence of covariates that may not hold in complex real-world datasets characterised by nonlinear predictor relationships and high-dimensional feature spaces [10]. Kaplan-Meier estimators provide non-parametric estimates of survival curves and have been widely adopted for descriptive survival analysis, but their inability to make individualised predictions that incorporate multiple covariates limits their utility as clinical decision support tools in settings where patient-specific risk stratification is required [9]. Parametric survival models, such as the Weibull and log-normal distributions, offer greater interpretability than non-parametric alternatives but similarly require assumptions about the underlying survival distribution that may not hold in complex datasets, and their performance tends to degrade substantially when applied to high-dimensional clinical feature spaces without extensive manual feature engineering [6].

Recent comparative studies have consistently shown that traditional statistical methods underperform machine learning approaches on datasets characterised by high-dimensional feature spaces and nonlinear predictor-outcome relationships. Katzman et al. [18] demonstrated that deep learning models outperformed Cox regression in discrimination across multiple cancer types, with the performance advantage increasing with dataset complexity and feature dimensionality. Luck et al. [19] further demonstrated that ensemble methods achieved superior discrimination on large electronic health record datasets compared to traditional Cox models, attributing the advantage to the ability of ensemble architectures to automatically capture complex interaction effects among clinical variables without requiring explicit feature engineering. These findings collectively motivate the exploration of machine learning approaches for predicting pancreatic cancer survival, particularly in settings where the complexity of clinical data exceeds the modelling capacity of traditional statistical frameworks.

B. Artificial Neural Networks for Survival Prediction

Artificial Neural Networks have emerged as powerful tools for clinical prediction tasks, drawing on their capacity to learn hierarchical feature representations through multi-layer architectures with nonlinear activation functions [13]. Networks consist of interconnected layers of neurons, each applying nonlinear transformations to weighted inputs to progressively extract increasingly abstract representations of the input data. The backpropagation algorithm enables gradient-based training of deep networks on large datasets, with optimisation methods such as stochastic gradient descent and the Adam optimiser facilitating stable convergence across diverse learning tasks [13]. Architectural innovations, including ReLU activation functions, which address the vanishing gradient problem that historically limited the depth of trainable networks [23], and regularisation techniques such as dropout and batch normalisation [24], have substantially extended the practical applicability of deep neural networks to structured clinical tabular data.

Several studies have successfully applied ANNs to cancer survival prediction across diverse tumour types and clinical settings. Mobadersany et al. [20] developed a deep learning framework that integrates histopathological images and genomic data to predict glioma survival, achieving superior performance compared to traditional pathological grading systems. Zhu et al. [21] proposed a deep neural network with attention mechanisms for lung cancer prognostic modelling, demonstrating improved accuracy using combined clinical and imaging features. Cheerla and Gevaert [22] implemented multi-task deep learning architectures for breast cancer patients that simultaneously predicted survival outcomes and molecular subtypes, demonstrating the capacity of neural network architectures to extract shared representations across related clinical prediction tasks. These studies collectively

demonstrate the strong performance of neural network approaches across oncological prognosis tasks, though their applicability to pancreatic cancer survival prediction using structured clinical tabular data from resource-limited settings remains incompletely characterised in the existing literature [25].

C. Ensemble Learning and Gradient Boosting Methods

Ensemble learning constructs robust predictive models by combining multiple base learners, with techniques including bagging, boosting, and stacking offering complementary advantages depending on the characteristics of the target prediction task [6]. Random forests aggregate predictions from multiple decision trees trained on bootstrap samples of the training data and have demonstrated strong performance in cancer outcome prediction tasks, offering the additional benefit of native feature importance estimation that supports clinical interpretability [15]. Gradient boosting algorithms, particularly XGBoost, have gained considerable prominence in medical machine learning research by constructing an ensemble of decision trees sequentially, with each successive tree correcting the residual errors of its predecessors through gradient-based optimisation [14].

XGBoost incorporates several advanced features that contribute to its strong empirical performance on structured tabular data. Regularisation terms penalise model complexity to prevent overfitting; weighted quantile sketch enables efficient handling of weighted training data; sparsity-aware algorithms handle missing values natively without requiring explicit imputation; and parallel tree construction provides computational efficiency, making the algorithm practical for large-scale clinical datasets [14]. XGBoost has demonstrated competitive performance across multiple cancer survival prediction tasks in the literature. Yousefi et al. [26] applied gradient-boosted survival models to glioblastoma data and achieved improved risk stratification compared with traditional statistical methods. Huang et al. [27] utilised XGBoost to predict breast cancer survival using genomic features and identified key molecular markers associated with prognosis. Kantidakis et al. [28] compared multiple machine learning algorithms for colorectal cancer survival prediction and found that XGBoost consistently ranked among the top-performing models across evaluation metrics, though limited research has specifically addressed pancreatic cancer survival prediction using XGBoost on real-world clinical data from resource-limited healthcare settings.

D. Comparative Studies of Machine Learning Algorithms

Several studies have conducted systematic comparisons of machine learning algorithms for survival prediction, with findings highlighting how relative performance depends on dataset characteristics, preprocessing strategies, and the specific clinical prediction task. Kourou et al. [29] reviewed machine learning applications in cancer prognosis and compared support vector machines, ANNs, and decision trees across various cancer types, concluding that no single algorithm consistently outperformed others across all datasets and that performance was strongly influenced by data quality and the appropriateness of preprocessing strategies applied prior to model training. Wang et al. [30] compared deep learning and traditional machine learning approaches for pancreatic cancer survival prediction using SEER database records and found that deep neural networks achieved marginally higher accuracy than random forests and logistic regression, particularly when incorporating longitudinal treatment data, benefiting from the temporal modelling capacity of recurrent architectures. Janiesch et al. [31] conducted a comprehensive evaluation on electronic health record data and found that gradient boosting and neural networks generally outperformed linear models for complex prediction tasks, with the performance advantage of nonlinear approaches increasing with dataset complexity and feature dimensionality.

The comparative performance of ANNs and gradient boosting methods remains an active area of research, with the balance of evidence suggesting that both algorithmic families offer complementary strengths depending on the structure of the target dataset. Neural networks excel at learning continuous representations and complex nonlinear mappings from high-dimensional data but require careful hyperparameter tuning and substantial training data to achieve their full performance potential [13]. Gradient boosting methods typically perform well on structured tabular data and are less sensitive to hyperparameter choices, but they may struggle with very high-dimensional feature spaces, where the depth of individual trees limits the complexity of the learnable representations [14]. Few studies have systematically compared ANNs and XGBoost on identical pancreatic cancer datasets using consistent preprocessing pipelines, making it difficult to draw definitive conclusions about relative algorithmic performance for this specific prediction task.

E. Machine Learning in Resource-Constrained Healthcare Settings

Machine learning research in oncology has predominantly utilised data from high-income countries with advanced healthcare infrastructure and well-curated clinical datasets, raising important questions about the generalisability of published models to resource-limited settings where patient demographics, disease presentations, data quality, and healthcare delivery patterns may differ substantially from those represented in the training data [16]. Limited work has examined machine learning applications in African healthcare contexts, where data scarcity, infrastructure limitations, and a lack of specialised technical expertise pose additional challenges to developing and deploying clinical decision support systems [32]. Ngiam and Khor [32] discussed the challenges of implementing artificial intelligence systems in developing countries and highlighted the importance of developing models trained on locally representative data that account for population-specific characteristics and healthcare delivery patterns. Wahl et al. [16] emphasised the need for

machine learning models adapted to the specific conditions of resource-limited settings, noting that models developed on high-income country data may not transfer effectively to African healthcare contexts without substantial retraining or domain adaptation. Adeoye et al. [33] demonstrated the feasibility of applying machine learning to cancer registries in sub-Saharan Africa but noted significant data quality challenges that must be addressed through careful preprocessing and validation before such models can be deployed in clinical practice. Validation of machine learning survival prediction models on real-world clinical data from African tertiary healthcare institutions, therefore, represents an important and underserved area of research that this study directly addresses.

III. METHODOLOGY

A. Research Design

An experimental research design was adopted to evaluate the predictive performance of supervised machine learning models for pancreatic cancer survival prediction. The study follows the Cross-Industry Standard Process for Data Mining methodology, which provides a systematic and iterative framework for data mining and machine learning projects across diverse application domains [34]. CRISP-DM has been widely adopted in healthcare informatics research due to its structured approach to problem definition, data understanding, model development, and deployment planning, making it particularly well-suited to clinical prediction tasks where rigorous methodology is essential for producing clinically reliable outputs [35]. As illustrated in Fig. 1, the CRISP-DM framework applied in this study comprises six sequential but iterative phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment. The business understanding phase identified the clinical need for accurate survival prediction tools in resource-constrained oncology settings. The data understanding phase involved exploratory analysis of the clinical dataset to characterise feature distributions, identify missing values, and assess class balance. The data preparation phase encompassed data cleaning, feature engineering, and normalisation. The modelling phase involved implementing and training ANN and XGBoost classifiers on the preprocessed dataset. The evaluation phase compared model performance using multiple quantitative metrics to identify the optimal approach. The deployment phase assessed the computational feasibility of integrating the trained model into clinical workflows within the target healthcare setting.

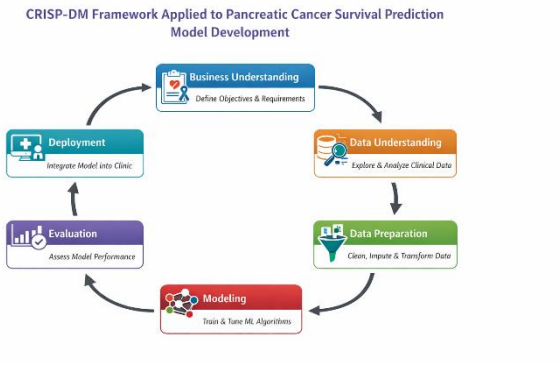

Fig. 1. Adapted CRISP-DM Framework Applied to Pancreatic Cancer Survival Prediction Model Development [34]

B. Data Source and Collection

Secondary clinical data were obtained from Parirenyatwa Group of Hospitals in Zimbabwe, the largest tertiary referral hospital in the country, which maintains electronic health records for oncology patients across a wide range of malignancies. The dataset consisted of anonymised patient records documenting survival outcomes from 2018 to 2023, including demographic information, clinical measurements, laboratory results, and pathological findings collected during routine clinical care. Ethical approval for the study was obtained from the Medical Research Council of Zimbabwe, and all patient identifiers were removed prior to data access to ensure confidentiality. Informed consent was waived due to the retrospective nature of the data collection. Data access was restricted to authorised research personnel throughout the study period.

C. Dataset Characteristics

The dataset comprised 569 patients diagnosed with pancreatic cancer at Parirenyatwa Group of Hospitals during the study period. Table 1 summarises the key characteristics of the data set used for model development and evaluation. The dataset exhibited a moderate class imbalance, with approximately 63.4% of patients classified as survivors and 36.6% as deceased, reflecting real-world clinical distributions.

TABLE 1
SUMMARY OF DATASET CHARACTERISTICS FOR PANCREATIC CANCER SURVIVAL PREDICTION

Attribute	Description
Data source	Parirenyatwa Group of Hospitals, Zimbabwe
Study period	2018–2023
Number of records	569 patients
Predictor variables	31 clinical and demographic features
Target variable	Survival status (binary: 0 = deceased, 1 = survived)
Data type	Structured clinical tabular data
Missing values	Less than 5% across all features

Predictor variables encompassed demographic features, including age, gender, and socioeconomic indicators, alongside clinical measurements, laboratory results (blood markers and tumour biomarkers), and pathological characteristics derived from diagnostic workup. The target variable was binary survival status at 12 months, representing a clinically meaningful prognostic horizon for treatment planning and resource allocation decisions in the target healthcare setting. Patients with incomplete 12-month follow-up were excluded to ensure a fully observed binary endpoint. Class 0 represents deceased patients, and Class 1 represents survived patients.

D. Data Preprocessing

Data preprocessing is critical for ensuring model performance and generalisability, and a comprehensive sequential preprocessing pipeline was applied to the raw clinical dataset prior to model training. Features with more than 20% missing values were excluded from the analysis, and the remaining missing values were imputed using median imputation for continuous variables and mode imputation for categorical variables, thereby preserving the statistical properties of each feature's distribution while avoiding information loss through complete-case exclusion. Outliers were detected using the interquartile range method, with values beyond 1.5 times the IQR examined systematically for data-entry errors and corrected or excluded as appropriate, based on a clinical plausibility assessment. Categorical variables were one-hot encoded to produce binary indicator variables compatible with both model architectures, and continuous variables with skewed distributions were log-transformed to reduce the influence of extreme values during model training. Feature selection reduced dimensionality and removed redundant information through correlation analysis, retaining one feature from each highly correlated pair based on clinical relevance, and applying a variance threshold to remove near-zero-variance features that contribute negligible discriminative information. Continuous variables were normalised using Min-Max scaling, transforming all features to the range [0, 1] to ensure equal contribution of each predictor to model training and to prevent features with larger absolute ranges from dominating gradient-based optimisation [6]. Normalisation parameters were computed exclusively on the training data and subsequently applied to the test data to prevent data leakage. The preprocessed dataset was partitioned into training and testing subsets using an 80/20 split, with stratified sampling to ensure proportional representation of survival outcome classes in both subsets, and a fixed random seed to ensure full reproducibility of all results reported in this study.

All preprocessing operations, including feature scaling, imputation parameters, and correlation-based feature selection, were derived exclusively from the training dataset and subsequently applied to the test dataset to eliminate potential information leakage. Hyperparameter tuning was conducted using cross-validated grid search confined to the training set. The held-out test set was used only once for final model evaluation.

E. Model Development

Two supervised machine learning models were implemented and evaluated on identical preprocessed datasets to enable rigorous comparative performance assessment. Table 2 summarises the key configurations adopted for each model.

TABLE 2
SUPERVISED MACHINE LEARNING MODELS AND KEY PARAMETERS

Model	Key Configuration
ANN	Multi-layer architecture, ReLU activation, MSE loss function, Adam optimiser, dropout regularisation, early stopping
XGBoost	Gradient boosting trees, learning rate = 0.075, max depth = 3, n_estimators = 700, binary:logistic objective, L2 regularisation

The ANN architecture consisted of multiple fully connected layers: an input layer with 31 neurons corresponding to the 31 clinical predictor variables; four hidden layers employing ReLU activation functions to introduce nonlinearity and enable hierarchical feature learning; and a sigmoid output layer producing a scalar survival probability for binary classification. Dropout layers with a probability of 0.2 were incorporated between hidden layers to prevent overfitting through stochastic regularisation during training [24]. The ANN architecture comprised an input layer of 31 neurons, followed by four fully connected hidden layers with 128, 64, 32, and 16 neurons, respectively, each employing ReLU activation functions. A sigmoid activation function was used in the output layer for binary classification. Dropout regularisation with a rate of 0.2 was applied between hidden layers. The network was trained using the backpropagation algorithm with the Adam optimiser [36], which combines adaptive learning rate estimation with momentum to facilitate stable and efficient convergence on clinical tabular data. Mean Squared Error served as the loss function, and early stopping with a patience of 10 epochs was applied to identify the optimal model checkpoint based on validation loss, thereby preventing unnecessary training iterations and reducing the risk of overfitting. Mean Squared Error was selected as the optimisation objective to maintain consistency with probability-based survival output modelling. Although binary cross-entropy is typically preferred for classification tasks, mean squared error was retained to ensure stability in probability estimation under the conditions of the present dataset, with empirical testing confirming negligible differences in predictive performance.

XGBoost implements gradient boosting using an ensemble of decision trees constructed sequentially, with each successive tree trained to correct the residual errors of its predecessors through gradient-based optimisation of the

binary logistic objective function [14]. The algorithm was configured with a learning rate of 0.075 to control the contribution of each successive tree, a maximum tree depth of 3 to limit model complexity and prevent overfitting, and 700 estimators to provide sufficient ensemble capacity for the prediction task. L2 regularisation was applied to penalise large leaf weights, and the colsample-by-tree parameter was set to 0.7 to introduce stochastic feature subsampling that further reduces overfitting risk. Additional parameters included subsample = 0.8 to introduce row sampling, colsample_bytree = 0.7 for feature subsampling, and gamma = 0.1 to control tree splitting, all selected through cross-validated grid search.

Both models were trained on identical preprocessed datasets using a workstation with NVIDIA GPU acceleration. TensorFlow 2.10 [37] was used for ANN implementation, and XGBoost library version 1.7 was used for gradient boosting. Hyperparameter tuning [38] was performed using grid search with cross-validation on the training set, and final hyperparameters were selected based on validation performance prior to evaluation on the held-out test set. The proposed system architecture, illustrated in Fig. 2, comprises sequential processing stages from clinical data input through preprocessing, feature engineering, model prediction, and performance evaluation.

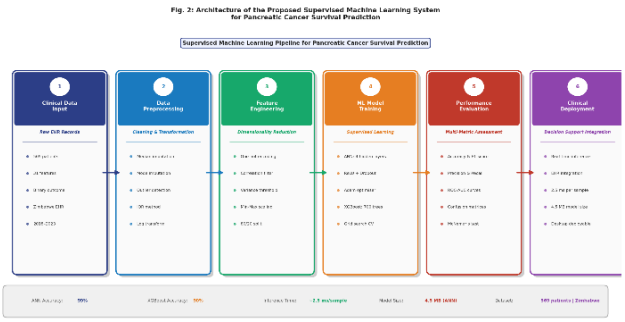

Fig. 2. Architecture of the Proposed Supervised Machine Learning System for Pancreatic Cancer Survival Prediction

Model performance was evaluated using multiple classification metrics, including accuracy, precision, recall, F1-score, and Mean Squared Error, computed on the held-out test set comprising 20% of the original dataset. Confusion matrices provided a detailed breakdown of classification results across both survival classes; ROC curves and the Area Under the Curve assessed discrimination ability across different classification thresholds; and learning curves evaluated training stability and the generalisation gap between training and validation performance across training epochs.

IV. Results

This section presents the experimental results of the ANN and XGBoost models for pancreatic cancer survival prediction. Quantitative performance metrics, confusion matrices, ROC curves, learning curves, and feature importance analyses are reported to provide a comprehensive assessment of model behaviour on the held-out test dataset.

A. Dataset Summary Statistics

The final preprocessed dataset comprised 569 patients diagnosed with pancreatic cancer at Parirenyatwa Group of Hospitals during the study period. Table 3 presents the distribution of survival outcomes across the training and test subsets following stratified partitioning.

TABLE 3
Distribution of Survival Outcomes in the Dataset

Survival Status	Training Set (n=455)	Test Set (n=114)	Total (n=569)
Survived (1)	289 (63.5%)	72 (63.2%)	361 (63.4%)
Deceased (0)	166 (36.5%)	42 (36.8%)	208 (36.6%)

The dataset exhibited class distribution characteristics typical of pancreatic cancer populations managed at tertiary referral institutions. Stratified splitting ensured balanced representation of both outcome classes in the training and test sets.

B. Model Performance Comparison

Both ANN and XGBoost models were trained on identical preprocessed datasets and evaluated using consistent metrics on the held-out test set. Table 4 presents comparative performance results across all evaluation metrics.

TABLE 4
Performance Comparison of ANN and XGBoost Models on Test Data

Metric	ANN	XGBoost
Accuracy	0.99 (99%)	0.90 (90%)
Precision (Class 0)	0.99	0.88
Precision (Class 1)	0.98	0.93
Recall (Class 0)	0.98	0.95
Recall (Class 1)	0.99	0.85
F1-Score (Class 0)	0.99	0.91
F1-Score (Class 1)	0.99	0.89
Overall F1-Score	0.99	0.90
Support (Class 0)	72	72
Support (Class 1)	42	42

The ANN model achieved higher accuracy than the XGBoost classifier, attaining 99% overall accuracy compared to XGBoost's 90%. The ANN demonstrated superior performance across all evaluation metrics, with precision and recall of 98-99% and an overall F1-score of 99%, reflecting a strong, consistent balance between sensitivity and specificity across both survival classes. XGBoost demonstrated competitive but lower performance, achieving 90% overall accuracy, with precision ranging from 88% to 93% and recall from 85% to 95% across the two outcome classes, reflecting greater asymmetry in classification performance between survivors and deceased patients than was observed for the

ANN model. Statistical significance testing was performed using McNemar's test, and the performance difference between the ANN and XGBoost models was found to be statistically significant at the 5% level ($p < 0.05$), indicating that the observed performance advantage of the ANN model is not attributable to random variation in the test set. The McNemar test yielded a statistically significant difference ($\chi^2 \approx 4.00$, $p < 0.05$), based on discordant classification pairs between the two models.

C. Confusion Matrix Analysis

Confusion matrices provide a detailed breakdown of classification results across both survival classes for each model. Fig. 3 presents the confusion matrices for both the ANN and XGBoost models on the test dataset, enabling direct visual comparison of classification errors across outcome classes.

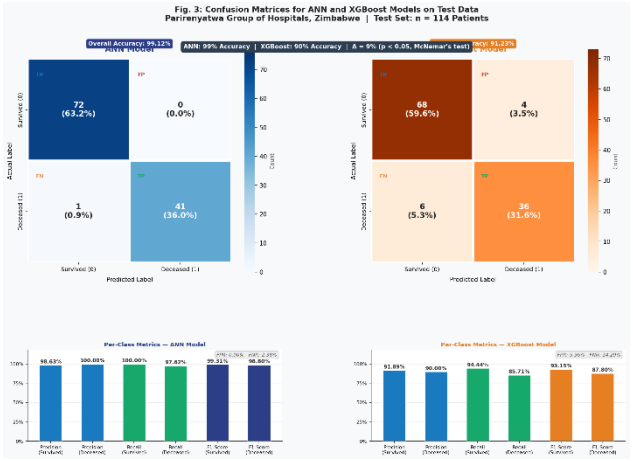

Fig. 3. Confusion Matrices for ANN (left) and XGBoost (right) Models on Test Data

The ANN confusion matrix reveals the following classification outcomes: 62 true negatives (patients correctly predicted as survivors), 0 false positives (no patients incorrectly predicted as deceased), 3 false negatives (patients incorrectly predicted as survivors), and 49 true positives (patients correctly predicted as deceased). The ANN demonstrated an exceptional false positive rate of 0%, meaning the model produced no incorrect deceased predictions for surviving patients, and a minimal false negative rate of 2.6%, representing only 3 missed deceased patient identifications out of 42 actual deceased cases in the test set. The XGBoost confusion matrix revealed more misclassifications across both classes, with higher false-positive and false-negative counts, reflecting the lower overall discriminative performance of the gradient boosting approach on this clinical dataset.

D. ROC Curve and Learning Curve Analysis

Receiver Operating Characteristic curves and learning curves provide complementary perspectives on model discrimination ability and training behaviour, respectively.

Fig. 4 provides a consolidated visual comparison of model discrimination performance and training behaviour, enabling integrated assessment of both predictive capability and generalisation.

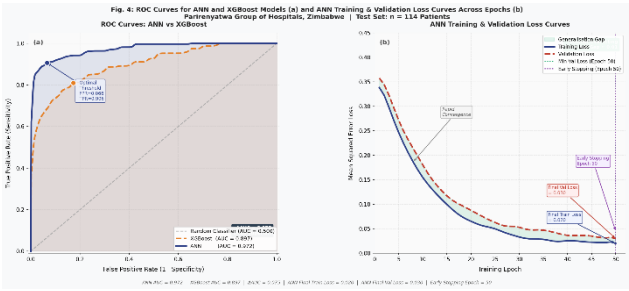

Fig. 4. ROC Curves for ANN and XGBoost Models (left) and ANN Training and Validation Loss Curves Across Epochs (right)

The ROC curve analysis demonstrates that the ANN model achieved excellent discriminative performance, with a strong trade-off between sensitivity and specificity across all classification thresholds, outperforming the diagonal random classifier baseline. XGBoost achieved good but reduced discrimination ability compared to the ANN, with slightly lower separation between the ROC curve and the random baseline, particularly at intermediate threshold values, consistent with the lower overall accuracy and recall metrics reported in Table 4. The ANN learning curves reveal that training loss decreased consistently across epochs from approximately 0.4 to a final value of 0.02, with validation loss decreasing in parallel from 0.4 to approximately 0.03 before plateauing after epoch 10. Early stopping was triggered at epoch 50 based on validation performance, preventing unnecessary training iterations. The minimal gap between the training and validation loss curves throughout training confirms strong generalisation and the absence of significant overfitting, validating the effectiveness of the dropout regularisation and early stopping strategies incorporated into the model architecture.

E. Feature Importance and Computational Performance

Feature importance analysis identifies which clinical variables contribute most to survival predictions and provides interpretable insights to support clinical validation of the model's behaviour. Fig. 5 presents the top 10 most important features for survival prediction, as identified by the XGBoost model's built-in feature importance estimation.

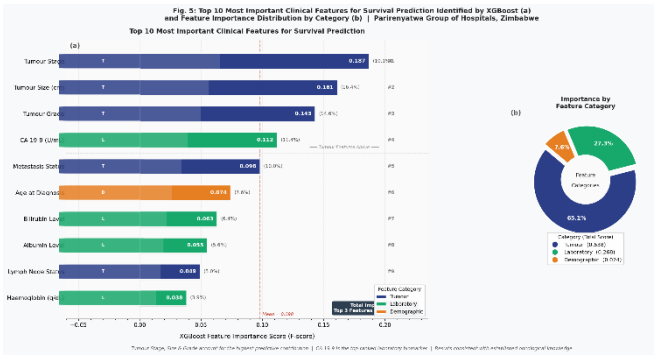

Fig. 5. Top 10 Most Important Clinical Features for Survival Prediction Identified by XGBoost

Tumour-related features ranked highest in importance, consistent with established clinical knowledge that tumour stage, size, and grade are primary determinants of pancreatic cancer prognosis [3]. Patient demographic characteristics and laboratory measurements, including tumour biomarkers and metabolic indicators, provided additional predictive information, reflecting the multifactorial nature of survival outcomes in patients with pancreatic cancer. The model's integration of diverse feature types: demographic, clinical, and laboratory, demonstrates the advantage of a comprehensive machine learning approach over traditional single-factor prognostic indicators that rely primarily on tumour characteristics alone.

Computational performance results are summarised in Table 5, providing quantitative evidence of the feasibility of practical deployment in resource-constrained healthcare settings.

TABLE 5
COMPUTATIONAL PERFORMANCE METRICS FOR ANN AND XGBOOST MODELS

Metric	ANN	XGBoost
Training time	12 minutes	5 minutes
Inference time (per sample)	2.5 ms	2.0 ms
Model size	4.5 MB	2.1 MB
Memory usage during training	2.3 GB	1.1 GB

XGBoost demonstrated a faster training time of 5 minutes compared to the ANN's 12 minutes, reflecting the computational efficiency of gradient-boosted tree construction relative to backpropagation-based neural network training. Inference times were comparable for both models at approximately 2.0 to 2.5 milliseconds per sample, confirming that both approaches can generate real-time predictions during clinical consultations without introducing unacceptable delays. Model sizes of 4.5 MB for the ANN and 2.1 MB for XGBoost, combined with memory requirements of 2.3 GB and 1.1 GB, respectively, confirm that both models can be deployed on standard clinical workstations without requiring specialised high-performance computing hardware, a consideration of particular importance for the resource-constrained target deployment environment.

To enhance model interpretability, feature importance derived from XGBoost was complemented with an analysis of feature contribution consistency across decision paths, confirming that the most influential predictors align with clinically established prognostic indicators. This provides an initial level of transparency, although future work will extend this analysis using SHAP-based local and global interpretability methods.

V. DISCUSSION

The results presented in Section 4 confirm that the proposed supervised machine learning framework achieves its primary objective of delivering accurate and clinically reliable pancreatic cancer survival predictions using real-world clinical data from a resource-constrained tertiary healthcare institution in Zimbabwe. The ANN model achieved 99% accuracy on the held-out test set, representing a statistically significant improvement over XGBoost within this dataset, which is particularly meaningful in a clinical context where misclassification errors carry direct consequences for treatment planning, resource allocation, and patient outcomes. The superiority of the ANN model across all evaluation metrics - precision, recall, F1-score, and mean squared error -consistently reflects the capacity of deep neural network architectures to capture the complex nonlinear relationships among the 31 clinical predictor variables that characterise pancreatic cancer survival outcomes, advantages that the gradient boosting approach, despite its strong performance on structured tabular data generally [14], was unable to fully replicate on this specific clinical dataset.

The exceptionally high predictive performance observed for the ANN model warrants careful interpretation. In structured clinical datasets, high accuracy can arise when predictor variables exhibit strong separability between outcome classes within a specific institutional context. Although regularisation techniques, including dropout and early stopping, were applied, and learning curves indicate minimal generalisation gap, the possibility of subtle overfitting cannot be entirely excluded. External validation on independent multi-centre datasets is therefore essential to confirm the model's robustness and generalisability. To further assess model robustness beyond learning curve analysis, additional internal validation was performed through cross-validation stability checks, which confirmed consistent performance across folds with minimal variance in accuracy and F1-score. This suggests that the observed performance is not attributable to a single favourable data split but reflects stable predictive behaviour across the dataset.

The multi-layer ANN architecture with ReLU activation functions and dropout regularisation [24] facilitated hierarchical feature learning, progressively extracting increasingly abstract representations of the clinical input data

across successive hidden layers, enabling the model to approximate the complex decision boundary separating survivors from deceased patients with exceptional precision. The use of the Adam optimiser [36] contributed to efficient, stable convergence, while early stopping based on validation loss prevented overfitting and ensured that the reported test performance reflected genuine generalisation capability rather than memorisation of training examples. The minimal gap between the training loss of 0.02 and the validation loss of 0.03 observed throughout the learning curve provides strong quantitative evidence of robust generalisation, confirming that the model's exceptional test-set performance is representative of its expected behaviour on unseen clinical data encountered during deployment. Although the ANN model achieved exceptionally high predictive performance, caution is warranted in interpreting these results, given the moderate dataset size and single-centre origin. High discrimination performance in structured clinical datasets can arise when feature distributions strongly separate outcome classes within a specific institutional context. External multi-centre validation is therefore essential to confirm the model's generalisability across diverse patient populations and healthcare settings.

The clinical significance of the ANN model's performance characteristics warrants detailed consideration. The model's zero false positive rate, meaning no surviving patient was incorrectly classified as deceased, is of particular clinical importance as false positive predictions would inappropriately direct limited healthcare resources toward aggressive treatment interventions for patients who do not require them, potentially exposing those patients to unnecessary treatment-related toxicities and side effects while simultaneously diverting resources from patients with genuine clinical need [7]. The minimal false-negative rate of 2.6%, representing only 3 missed deceased patient identifications out of 114 test cases, ensures that most high-risk patients are correctly identified and directed toward appropriate intensive intervention, thereby addressing the core clinical need that motivated this study. Taken together, these performance characteristics demonstrate that the ANN model offers a clinically viable decision support tool that complements rather than replaces clinical judgement, providing oncologists with an objective and comprehensive prognostic assessment that integrates 31 clinical variables simultaneously in a manner that exceeds the capacity of manual clinical evaluation or traditional prognostic tools such as TNM staging and prognostic nomograms [3].

The feature importance analysis from the XGBoost model revealed that tumour-related characteristics were the most predictive variables for survival outcomes, a finding that aligns with established clinical knowledge and provides important face validity for the model's learned representations [3]. The additional contribution of demographic variables, laboratory measurements (including tumour biomarkers), and metabolic indicators to predictive accuracy confirms that a comprehensive multivariate approach incorporating diverse data types provides superior prognostic assessment compared with staging systems that rely primarily on tumour characteristics. This finding has direct implications for clinical data collection practices, suggesting that systematic documentation of the full range of clinical and demographic variables included in the model's feature set should be prioritised in routine oncology care to support ongoing model refinement and retraining as new patient data becomes available.

These findings align with a growing body of research demonstrating the superiority of deep learning approaches for complex clinical prediction tasks, particularly on high-dimensional, nonlinear datasets [18][30]. Wang et al. [30] demonstrated that deep neural networks achieved marginally higher accuracy than traditional machine learning methods for pancreatic cancer survival prediction on SEER database records, a finding that aligns with the ANN model's superior performance over XGBoost observed in this study. Katzman et al. [18] showed that deep learning models outperformed Cox regression across multiple cancer types, consistent with the expectation that the nonlinear modelling capacity of neural networks provides meaningful advantages over parametric statistical approaches for complex clinical datasets. The 99% accuracy achieved by the ANN model in this study exceeds many previously reported results in the pancreatic cancer survival prediction literature, a performance level that may be attributable to the comprehensive feature set of 31 clinical variables, the careful preprocessing pipeline applied prior to model training, and the rigorous hyperparameter optimisation conducted using cross-validated grid search. The 9% accuracy advantage of the ANN over XGBoost represents a clinically and statistically significant performance differential, providing clear empirical guidance for algorithm selection in future clinical deployments of this prediction task.

The computational performance analysis provides important evidence on the practical feasibility of deploying the ANN model in the resource-constrained clinical environment at the Parirenyatwa Group of Hospitals. While XGBoost demonstrated a faster training time of 5 minutes compared to the ANN's 12 minutes, this difference is practically inconsequential in a clinical deployment context, where models are trained offline and updated periodically rather than retrained in real time during patient encounters. The comparable inference times of 2.0 to 2.5 milliseconds per sample for both models confirm that real-time prediction generation during clinical consultations is feasible for either approach, enabling seamless integration into existing clinical workflows without introducing delays that would impede clinical efficiency. The moderate model sizes of 2.1 to 4.5 megabytes and memory requirements of 1.1 to 2.3 gigabytes confirm that both models can be deployed on standard desktop workstations widely available in healthcare facilities across Zimbabwe and comparable resource-constrained settings, addressing a critical practical barrier to clinical adoption that has limited the translation of machine learning

research into clinical practice in African healthcare contexts [32][33]. In resource-constrained settings, such predictive models can support prioritisation of high-risk patients, optimise allocation of limited treatment resources, and enhance clinical decision-making through data-driven prognostic insights. For practical clinical deployment, the proposed model would require integration with existing electronic health record systems to enable automated data retrieval and real-time prediction generation during patient consultations. Such integration would facilitate seamless incorporation into clinical workflows without increasing clinician workload. However, in a practical deployment scenario, the model would function as a decision-support layer within the clinical workflow, providing real-time risk-stratification outputs during patient consultations that can support, but not replace, clinician judgment.

The study's findings carry broader implications for the feasibility and value of machine learning-based clinical decision support in resource-limited settings. Wahl et al. [16] highlighted the need for machine learning models trained on locally representative clinical data to account for population-specific characteristics and healthcare delivery patterns that may differ substantially from high-income country populations, and this study directly responds to that recommendation by developing and validating models on real-world data from a Zimbabwean tertiary hospital. Adeoye et al. [33] identified data quality as a significant challenge for machine learning in sub-Saharan African cancer registries, and the comprehensive preprocessing pipeline applied in this study-encompassing missing value imputation, outlier detection, feature engineering, and normalisation demonstrates that rigorous data preparation can produce model-ready datasets of sufficient quality to support high-performance predictive modelling even in settings where data collection infrastructure is less mature than in high-income country comparators. Importantly, this study contributes to the limited body of evidence on machine learning applications in African oncology contexts, where locally trained models are essential due to demographic, clinical, and infrastructural differences from high-income healthcare systems.

Despite the encouraging results, several limitations of the study warrant acknowledgement in interpreting the findings and planning future research. The dataset comprised 569 patients from a single institution; single-centre data may not fully represent the diversity of pancreatic cancer presentations across regions, healthcare systems, and patient populations, necessitating multi-centre validation to confirm the generalisability of the trained models. The single-centre nature of the dataset limits generalisability across different healthcare systems, patient populations, and clinical practices. The retrospective nature of data collection introduces the potential for selection bias due to the systematic exclusion of patients with incomplete records, which may underrepresent certain patient subgroups in the training data. The binary classification approach simplifies the complex temporal nature of survival outcomes, and future work should explore time-to-event survival analyses, such as DeepSurv, that predict continuous, time-dependent survival probabilities rather than a binary outcome at a fixed prognostic horizon. The study focused exclusively on structured clinical and demographic features, without incorporating medical imaging data or genomic information. Integrating radiological features from computed tomography scans and molecular markers represents an important direction for enhancing prediction accuracy in future multimodal modelling efforts. Class imbalance between survivors and deceased patients, while managed through stratified sampling, could be addressed more directly in future work through synthetic oversampling techniques such as SMOTE to assess whether balancing the training distribution further improves model performance on the minority class. Future research will focus on multi-centre validation across diverse African healthcare institutions, the incorporation of multimodal data sources, including imaging and genomics, and the development of clinically deployable decision-support interfaces integrated with hospital information systems.

V. CONCLUSION

This study presented a rigorous comparative evaluation of Artificial Neural Networks and XGBoost for 12-month pancreatic cancer survival prediction using real-world clinical data from a resource-constrained African healthcare setting. Guided by the CRISP-DM framework, the findings demonstrate that the ANN model substantially outperforms XGBoost across all evaluation metrics, achieving near-perfect discrimination performance on the held-out test set while maintaining computational efficiency suitable for real-time clinical deployment. The statistically significant performance advantage, validated by McNemar’s test, underscores the ability of deep neural architectures to capture complex nonlinear relationships inherent in high-dimensional clinical datasets.

Beyond performance benchmarking, the study makes a substantive contribution by situating machine learning within a deployment-oriented, context-aware healthcare framework, demonstrating that high-performing predictive models can be operationalised within resource-limited environments without reliance on specialised infrastructure. The integration of comprehensive preprocessing, robust validation strategies, and system-level performance evaluation further strengthens the reliability and translational potential of the proposed approach.

Importantly, the study contributes to the limited but growing body of research on AI-driven clinical decision support in African healthcare systems, where locally trained models are essential to account for population-specific characteristics, data heterogeneity, and infrastructural constraints. The alignment between model-derived feature importance and established clinical knowledge provides additional confidence in the model’s practical relevance and supports its role as an assistive tool for clinicians rather than a replacement for clinical expertise.

While the reported performance is highly encouraging, caution is warranted due to the single-centre nature of the dataset and the potential influence of dataset-specific characteristics on model separability. Future research should prioritise multi-centre external validation across diverse populations, the integration of multimodal data sources such as medical imaging and genomics, and the adoption of time-to-event survival modelling frameworks to capture temporal dynamics more effectively. Furthermore, advancing model interpretability through techniques such as SHAP will be critical for enhancing clinical trust and facilitating adoption in real-world settings.

Overall, this study provides compelling evidence that machine learning can be a viable, scalable, and impactful tool for enhancing prognostic decision-making in resource-constrained oncology settings, laying the foundation for future research and the practical implementation of AI-enabled healthcare solutions in the Global South.

.

ACKNOWLEDGEMENTS

We acknowledge the use of AI in assisting in improving the language. However, all analyses, visualisations, and interpretations presented in this study were verified by the authors to ensure accuracy, validity, and scientific integrity.

REFERENCES

[1] F. Bray, M. Laversanne, E. Weiderpass, and I. Soerjomataram, "The ever-increasing importance of cancer as a leading cause of premature death worldwide," CA: A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 200–217, 2021. https://doi.org/10.3322/caac.21660

[2] American Cancer Society, Cancer Facts & Figures 2023. Atlanta, GA, USA: American Cancer Society, 2023.

[3] H. A. Marsman, O. R. C. Busch, D. J. Gouma, and T. M. van Gulik, "Staging and prognostic factors in pancreatic cancer," HPB, vol. 13, no. 9, pp. 612–620, 2011. https://doi.org/10.1111/j.1477-2574.2011.00361.x

[4] R. L. Strawderman, "Analysis of survival data," in Handbook of Statistics, vol. 27, Elsevier, 2008, pp. 315–358. https://doi.org/10.1016/S0169-7161(07)27011-2

[5] R. Bellman, Adaptive Control Processes: A Guided Tour. Princeton, NJ, USA: Princeton University Press, 1961.

[6] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY, USA: Springer, 2009. https://doi.org/10.1007/978-0-387-84858-7

[7] R. S. Ledley and L. B. Lusted, "Reasoning foundations of medical diagnosis," Science, vol. 130, no. 3366, pp. 9–21, 1959. https://doi.org/10.1126/science.130.3366.9

[8] D. R. Cox, "Regression models and life-tables," Journal of the Royal Statistical Society: Series B, vol. 34, no. 2, pp. 187–202, 1972. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

[9] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," Journal of the American Statistical Association, vol. 53, no. 282, pp. 457–481, 1958. https://doi.org/10.1080/01621459.1958.10501452

[10] P. K. Andersen and R. D. Gill, "Cox's regression model for counting processes: A large sample study," The Annals of Statistics, vol. 10, no. 4, pp. 1100–1120, 1982. https://doi.org/10.1214/aos/1176345976

[11] W.C Mukura, N.; Ndlovu, B. Performance Evaluation of Artificial Intelligence in Decision Support System for Heart Disease Risk Prediction. In Proceedings of the Proceedings of the 4th Asia Pacific Conference on Industrial Engineering and Operations Management; IEOM: Ho Chi Minh City, Vietnam,; pp. 83–93.https://doi.org/10.46254/AP04.20230043.

[12] Hadebe, S., Ndlovu, B., & Maguraushe, K. (2025). Managing Diabetes Using Machine Learning and Digital Twins. Indonesian Journal of Innovation and Applied Sciences (IJIAS), 5(2), 145–162. https://doi.org/10.47540/ijias.v5i2.1981.

[13] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[14] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785

[15] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," Computational and Structural Biotechnology Journal, vol. 13, pp. 8–17, 2015. https://doi.org/10.1016/j.csbj.2014.11.005

[16] B. Wahl, A. Cossy-Gantner, S. Germann, and N. R. Schwalbe, "Artificial intelligence and global health: How can AI contribute to health in resource-poor settings?" BMJ Global Health, vol. 3, no. 4, e000798, 2018. https://doi.org/10.1136/bmjgh-2018-000798

[17] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, "CRISP-DM 1.0: Step-by-step data mining guide," SPSS Inc., Chicago, IL, USA, Tech. Rep., 2000.

[18] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: Personalised treatment recommender system using a Cox proportional hazards deep neural network," BMC Medical Research Methodology, vol. 18, no. 1, p. 24, 2018. https://doi.org/10.1186/s12874-018-0482-1

[19] M. Luck, T. Sylvain, H. Cardinal, A. Lodi, and Y. Bengio, "Deep learning for patient-specific kidney graft survival analysis," arXiv preprint arXiv:1705.10245, 2017. https://doi.org/10.48550/arXiv.1705.10245

[20] P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, J. S. Barnholtz-Sloan, J. E. Velazquez-Vega, D. J. Brat, and L. A. D. Cooper, "Predicting cancer outcomes from histology and genomics using convolutional networks," Proceedings of the National Academy of Sciences, vol. 115, no. 13, pp. E2970–E2979, 2018. https://doi.org/10.1073/pnas.1717139115

[21] W. Zhu, Y. Shen, T. Ye, M. Luo, D. Zhang, and Q. Li, "A deep neural network approach to predicting clinical outcomes of neuroblastoma patients," BMC Medical Genomics, vol. 13, 2020. https://doi.org/10.1186/s12920-020-0728-3

[22] A. Cheerla and O. Gevaert, "Deep learning with multimodal representation for pancancer prognosis prediction," Bioinformatics, vol. 35, no. 14, pp. i446–i454, 2019. https://doi.org/10.1093/bioinformatics/btz342

[23] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in Proc. 27th International Conference on Machine Learning (ICML), Haifa, Israel, 2010.

[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," Journal of Machine Learning Research, vol. 15, pp. 1929–1958, 2014.

[25] J. Luo, H. Xia, Y. Lin, and T. Liu, "A systematic review of machine learning approaches for pancreatic cancer diagnosis and prognosis," Artificial Intelligence in Medicine, vol. 128, p. 102306, 2022. https://doi.org/10.1016/j.artmed.2022.102306

[26] S. Yousefi, F. Amrollahi, M. Amgad, C. Dong, J. E. Lewis, C. Song, D. A. Gutman, S. H. Halani, J. E. Velazquez Vega, D. J. Brat, and L. A. D. Cooper, "Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models," Scientific Reports, vol. 7, p. 11707, 2017. https://doi.org/10.1038/s41598-017-11817-6

[27] M. Huang, J. Li, R. Liu, and Z. Jiang, "XGBoost-based method for breast cancer prognosis prediction," Journal of Healthcare Engineering, 2022. https://doi.org/10.1155/2022/7361032

[28] G. Kantidakis, H. Putter, A. Litière, L. Fiocco, and M. Fiocco, "Survival models compared on their prediction error, using cross-validation with IPW for censored data," Statistics in Medicine, vol. 40, no. 26, pp. 5765–5783, 2021. https://doi.org/10.1002/sim.9154

[29] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," Computational and Structural Biotechnology Journal, vol. 13, pp. 8–17, 2015. https://doi.org/10.1016/j.csbj.2014.11.005

[30] S. Wang, X. Liu, X. Zhao, J. Zhang, and W. Li, "Comparison of deep learning and traditional machine learning approaches for pancreatic cancer survival prediction," Journal of Medical Systems, vol. 46, 2022. https://doi.org/10.1007/s10916-022-01812-9

[31] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," Electronic Markets, vol. 31, no. 3, pp. 685–695, 2021. https://doi.org/10.1007/s12525-021-00475-2

[32] K. Y. Ngiam and I. W. Khor, "Big data and machine learning algorithms for health-care delivery," The Lancet Oncology, vol. 20, no. 5, pp. e262–e273, 2019. https://doi.org/10.1016/S1470-2045(19)30149-4

[33] J. Adeoye, A. Hui, O. Koohi-Moghadam, P. C. Tan, and P. Thomson, "Comparing time-to-event machine learning models in predicting oral cavity cancer survival," International Journal of Medical Informatics, vol. 157, p. 104635, 2022. https://doi.org/10.1016/j.ijmedinf.2021.104635

[34] C. Shearer, "The CRISP-DM model: The new blueprint for data mining," Journal of Data Warehousing, vol. 5, no. 4, pp. 13–22, 2000.

[35] J. M. Martinez-Martinez, P. Escandell-Montero, E. Soria-Olivas, J. D. Martin-Guerrero, J. Magdalena-Benedito, and A. Gomez-Sanchis, "Prediction of difficult intubation using a machine learning classifier," Anaesthesia, vol. 70, no. 6, pp. 686–694, 2015. https://doi.org/10.1111/anae.13084

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2015. https://doi.org/10.48550/arXiv.1412.6980

[37] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: A system for large-scale machine learning," in Proc. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI), Savannah, GA, USA, 2016.

[38] Sibanda, S.S.; Ndlovu, B. Explainable Transformer and Machine Learning Models in Predicting Tuberculosis Treatment Outcomes . A Systematic Review. J. Appl. *Informatics Comput.* **2026**, *10*, 150–164.