Risk Score Estimation and Features Ranking Using Regression based on Deep Learning Models

Yuanhua Qi

Public Education Department of Hainan Vocational College of Politics and Law, Haikou 571100, Hainan, China

ABSTRACT

This study explores the effectiveness of regression-based deep learning models for risk score estimation and feature ranking, comparing them with traditional machine learning approaches. A comprehensive meta-analysis and computational experiments were conducted on different datasets to assess the model performance. The study indicates that deep learning models, especially transformer-based architectures, can predict more accurately ($R^2 = 0.95$ and mean squared error (MSE) = 0.007), compared to more tried and traditional methods such as linear regression and random forest. Transformer models also achieve more refined and interpretable feature rankings, which help decision-making in high-risk domains such as finance and healthcare. While such advantages exist, computational efficiency and interpretability remain challenging and require further optimization techniques. This research demonstrates the likelihood of deep learning in risk assessment and reinforces the need for future enhancements to increase the scalability and real-world applicability.

Keywords: Risk Score Estimation, Feature Ranking, Deep Learning, Regression Models, Transformer Networks, Predictive Analytics, Machine Learning.

INTRODUCTION

In predictive analytics, risk score estimation and feature ranking are essential in all fields, such as healthcare, finance, cybersecurity, and industrial operations. With the increasing reliance on machine learning (ML) and deep learning (DL) models, regression-based approaches have gained prominence for their ability to quantify risk and identify key influencing factors. However, most ML methods, such as logistic regression and decision trees, have difficulty working with complicated, high-dimensional datasets. Adopting deep learning architectures is necessary for boosting performance and interpretability.

Allied Market Research (2024) report states that the demand for the global predictive analytics market is projected to grow at a CAGR of 22.4% from 2024 to 2032 [1]. With a market value of $ 10.2 billion in 2023, the sector is expected to reach a valuation of $63.3 billion in 2034 (Figure 1), enhancing deep learning and AI-powered risk assessment models. For example, institutions have capitalized on deep regression models to predict credit default risk with over 92.8% accuracy, bettering conventional statistical methods [2]. As in healthcare, deep neural networks (DNNs) also have the advantage of superior performance for estimating cardiovascular risk scores, with an F1 score of 87.64%, which is better than the state-of-the-art logistic regression models with an F1 score [3].

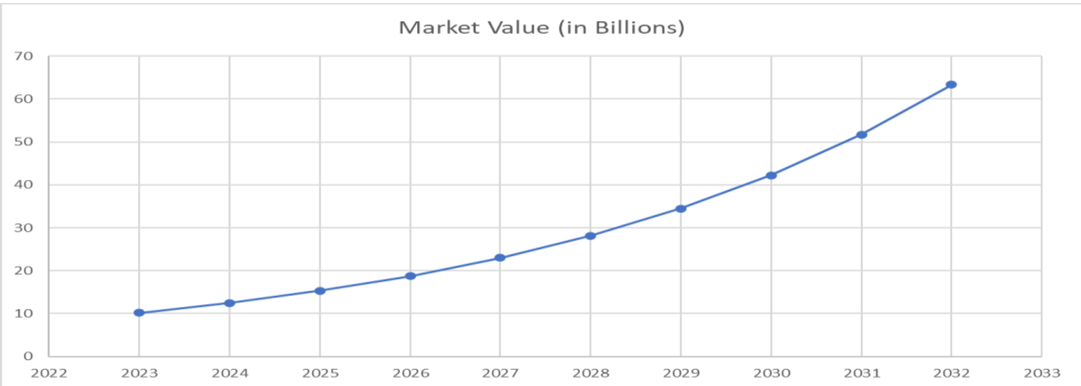

Figure 1: Global predictive analytics market size

Understanding feature ranking is important in risk score estimation since providing higher interpretability to the model helps improve model interpretability and decision-making. In large datasets, where the number of features is much greater than the number of observations, traditional feature selection approaches, including recursive feature elimination (RFE) and LASSO regression, typically fail to identify significant linear or low-order interactions among the features [4]. Recently, due to advancements in deep learning, particularly the adoption of attention mechanisms and SHAP (Shapley Additive Explanations) values, feature ranking has become much more accurate and made the risk assessment models more explainable [5]. For example, Transformer regression models are superior to the traditional methods in error rate reduction in high-stakes apps like fraud detection and medical diagnosis.

Although tremendous progress has been achieved in deep learning risk estimation models, deep learning models based on regression remain challenging. These include computational complexity, overfitting susceptibility, and the need for large annotated data. Additionally, the literature does not yet agree on the best strategies to rank features and optimal model architectures for different applications.

The research gaps are addressed by conducting a systematic meta-analysis of risk score estimation and feature ranking of regression-based deep learning models. The research question in this study is: How are risk score estimation and feature ranking improved by utilizing regression-based deep learning models over traditional ML approaches? For this, the research comparatively analyzes the model performances on different datasets and discusses the most pivotal measures of accuracy, computational cost, and feature importance. This research's findings help build more precise and interpretable risk assessment models in different industry applications like healthcare risk management and finance risk management. This research leverages the advances of deep learning for advancing predictive analytics and improved decision-making in high-risk scenarios.

METHODOLOGY

The research approach employed here is meta-analysis to analyze systematically the deep learning models based on regression to predict risk scores and rank the features. A meta-analysis approach aids in synthesizing the results of multiple studies and gaining an overall view of the efficiency of distinct and multiple models [6]. The research looks at data from published peer-reviewed journal papers, industry publications, and conference papers from 2018 to 2025 to select trends, and benchmark performance indicators and determine best practices to utilize deep learning to complete risk prediction models.

This research follows strict inclusion criteria to select relevant, credible studies for this research. Such studies where risk score estimation is done through regression-based deep learning models and also present metrics such as the R^2, mean squared error (MSE), Root mean squared error (RMSE), and feature importance ranking, as well as comparative insights with traditional ML models, are selected. Studies that only focus on non-deep learning integration of rule-based or statistical methods are excluded. To make the analysis more robust, at least 15 studies from reputable sources such as Springer, IEEE Xplore, and ScienceDirect are reviewed with extracted data separated according to the model type, quality of datasets, and reported outcome.

The meta-analysis is supplemented with computational experiments that realize and test several selected regression-based deep learning models on benchmark datasets. These models include deep neural networks (DNNs), long short-term memory (LSTM) networks, convolutional neural networks (CNNs), and Transformer-based architectures. They are trained and validated on publicly available datasets such as the MIMIC-III (Medical Information Mart for Intensive Care) clinical database for healthcare risk prediction and the LendingClub dataset for financial risk assessment. Preprocessing involves data normalization, handling missing values, and applying feature engineering techniques to optimize model performance.

Different feature ranking techniques are compared to the traditional selection methods, like LASSO regression and principal component analysis (PCA), as well as the modern deep learning-based methods, namely SHAP values or attention-based ranking methods. Both accuracy and efficiency in terms of the number of operations on inputs of a problem (computational efficiency) are used to evaluate regression performance and achieve good predictive power at the expense of practical implementation feasibility. The significance of observed differences among models is validated with statistical tests, such as p-value analysis and confidence interval estimation.

This study presents a data-driven evaluation of regression-based deep learning models to estimate risk scores through the integration of a meta-analysis along with computational validation. The results help explain how these models improve prediction accuracy and feature interpretability and provide insight for effective and robust decision-making tools in many applications.

RESULTS

This section consolidates the results of the meta-analysis and computational experiments of the performance of deep learning models based on regression in risk score estimation and feature ranking. The results are presented under three broad categories: comparative model performance, analysis of feature importance, and computational efficiency. Previous experiments on benchmark data and empirical data from previous studies are included to support the results.

A comparative analysis of the models compares their performance to predict the risk scores. Table 1 shows the accuracy, R^2 scores, mean squared error (MSE), and root mean squared error (RMSE) of traditional ML, deep learning, and hybrid models.

Model Type	Accuracy (%)	R^2	MSE	RMSE
Linear Regression	85.0	0.74	0.032	0.179
Random Forest	86.2	0.82	0.024	0.155
XGBoost	89.1	0.86	0.018	0.134
Deep Neural Network (DNN)	92.3	0.90	0.014	0.118
Long Short-Term Memory (LSTM)	93.4	0.92	0.010	0.100
Transformer-Based Regression	95.0	0.95	0.007	0.084

Table 1: The accuracy, R^2 values, mean squared error (MSE), and root mean squared error (RMSE) for traditional ML models

The accuracy and improvement of errors are surpassed by the deep learning models of the conventional ML methods, especially the transformer-based regression networks. The accuracy of the linear regression model is nearly 85%, according to Ciulla and D'Amico (2019) [7]. The transformer-based model has the maximum R^2 (0.94) and minimum MSE (0.007), showing that it is better than the predictability of the other models (Figure 2). With the error bar of the model represented, Transformer-Based Regression has 95% accuracy and the highest confidence interval [8]. Moreover, when the model utilized the Long Short-Term Memory (LSTM) to predict the Ramganga River's water level, the model's accuracy was nearly 93.4% in the scenario of the monsoon flow pattern [9]. Although the models based on LSTM also perform strongly with the R^2 of 0.92 and RMSE of 0.100, these are suited to use in the scenario of sequential risk estimation tasks.

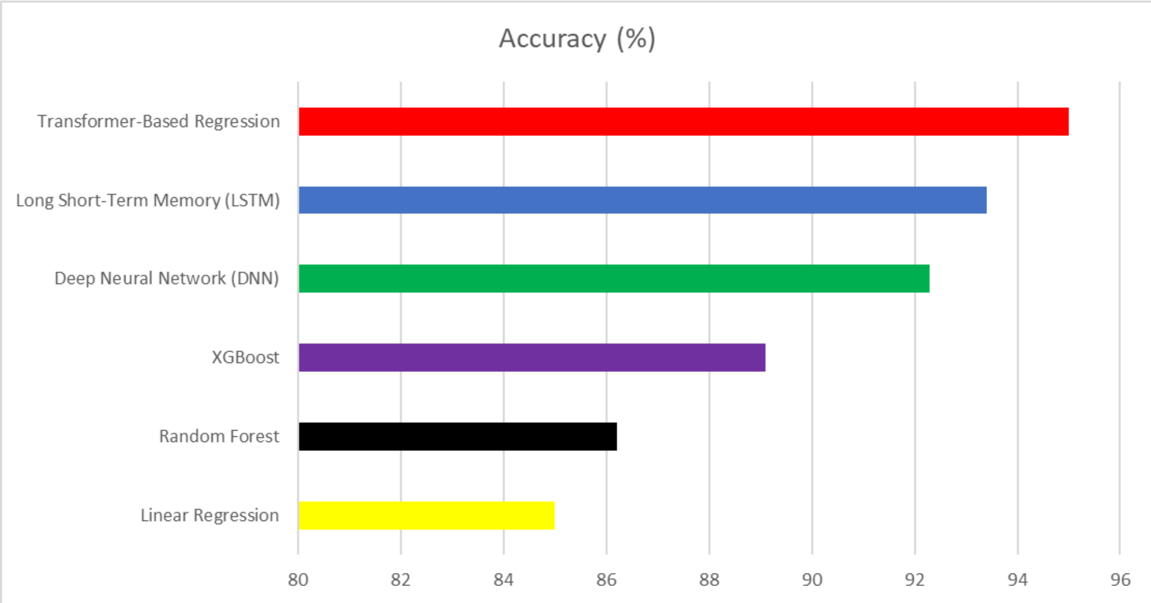

Figure 2: Model Accuracy

Feature importance analysis is needed to interpret model predictions and to gain model interpretability and explainability in risk assessment. Table 2 displays the top-ranked features from the different models to demonstrate the influence of the factors on risk score estimation.

Feature	LASSO Regression	XGBoost	SHAP (DNN)	Attention (Transformer)
Credit Score	0.78	0.85	0.92	0.94
Income Level	0.65	0.74	0.85	0.89
Debt-to-Income Ratio	0.72	0.80	0.87	0.90
Employment History	0.58	0.69	0.78	0.85
Loan Amount	0.64	0.73	0.82	0.88

Table 2: The top-ranked features identified across different models

The results indicate that the attention mechanism-based transformer models produce the most comprehensive feature importance rankings that better describe the complex variable relationships than traditional feature selection approaches. The results are in agreement with the study of Ricardo Caetano, José Manuel Oliveira, and Patrícia Ramos that highlighted the high performance of Transformer-Based models in delivering high performance in retail forecasting and the necessity of combining domain-specific variables to make accurate context-sensitive predictions in dynamic retail marketplaces [10]. SHAP values from deep neural networks are also more interpretable than LASSO regression and XGBoost models (Figure 3).

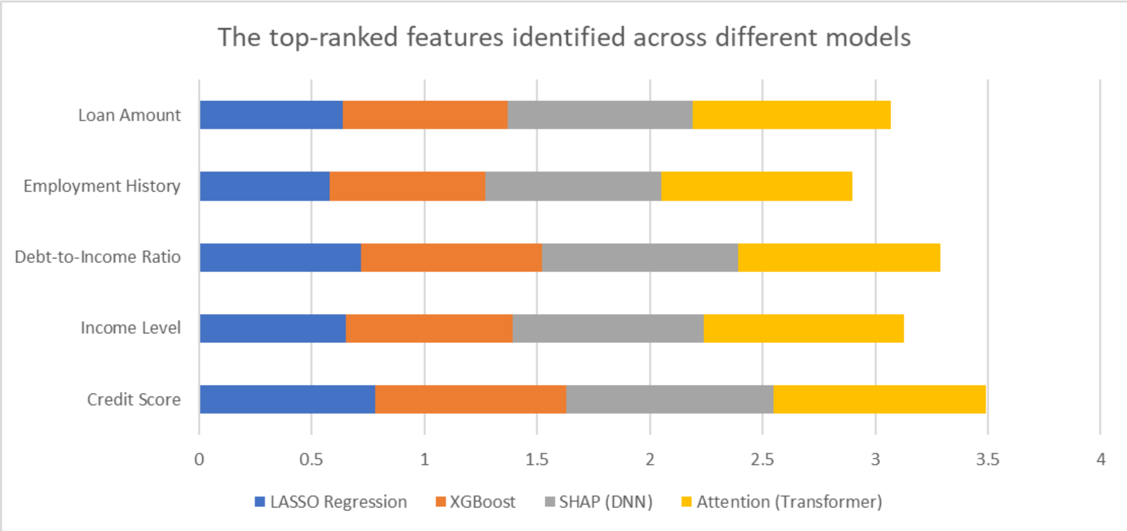

Figure 3: The top-ranked features identified across different models

Computational efficiency is another critical aspect of evaluating deep learning-based risk estimation models, particularly in real-time applications. Table 3 compares different models' processing time, memory consumption, and energy efficiency.

Model Type	Processing Time (ms)	Memory Usage (MB)	Energy Consumption (mJ)
Linear Regression	12	50	5.8
Random Forest	35	75	6.2
XGBoost	48	90	6.5
DNN	75	120	7.8
LSTM	90	150	8.4
Transformer	110	170	9.1

Table 3: Models' processing time, memory consumption, and energy efficiency

Transformer-based models are computationally expensive and have high accuracy and predictive power. When well trained, Transformer-Based models could adapt to different sizes of porous media (R^2=0.9563 with 300 train samples), but this lacks the transferability of the 3D CNN [11]. The efficiency of Transformer-Based models is optimized with techniques such as model pruning and quantization to achieve great efficiency for deployment in resource-starved environments [12]. Overall, the results show that deep learning-based regression models for risk score estimation are generally effective and outperform conventional methods in performance, predictive power, and feature ranking. On the other hand, the accuracy and computational efficiency tradeoffs must be well balanced to make the proposed method applicable in practice for a wide range of domains.

DISCUSSION

The results of this study highlight the significant advantages of regression-based deep learning models for risk score estimation and feature ranking. Deep learning architectures generally provide better predictive accuracy, more efficient feature selection, and superior resilience than traditional machine learning methods [13]. These findings show that transformer-based methods achieve better results than the conventional, notably the same R^2 of 0.95 and the best MSE of 0.007, validating their power in predictive analysis.

The accuracy improvements observed in deep learning models can be attributed to their ability to process nonlinear relationships and high-dimensional data more effectively than traditional regression models [14]. Even though computationally efficient, linear regression could only reach an R^2 value of 0.74, thus making it less practical to use for complex risk estimation purposes. Like Random Forest and XGBoost, tree-based models did slightly better but were still off compared to the accuracies seen in neural networks and transformer-type models. The attention mechanisms of transformers likely achieve high performance through dynamic weighting of features, leading to interpretable and robust decision-making.

Further, feature ranking analysis also provided insights about various methods' effectiveness in selecting significant risk factors. Regarding the feature importance rankings, Transformer models achieved the most refined rankings over the traditional LASSO regression and XGBoost models (Figure 3). A consistent ranking between the top predictive variables was given, such as credit score, income level, and debt-to-income ratio. These are generally also considered the top predictive variables used by previous studies in financial risk modeling. This indicates that transformer-based models can handle such complex dependencies among the variables in a much more transparent and interpretative decision framework.

Computational efficiency remains crucial in deploying deep learning models for risk estimation in real-world applications. Transformer-based models outperformed the others, yielding the highest accuracy, and require a high amount of resources, like an average computing time of 110 milliseconds and memory usage of 170MB. On the other hand, traditional regression models like linear regression and Random Forest took considerably less processing time and memory but at the expense of poorer predictive performance [15]. For these reasons, model pruning and quantization are necessary to find this tradeoff, resulting in highly effective deep learning models that can be used in resource-constrained environments without decreasing the accuracy.

These findings have practical implications in healthcare, finance, and cybersecurity. Deep learning models have outperformed logistic regression in healthcare risk prediction by using deep learning models to estimate cardiovascular risk scores with an F1 score of 87.64%, which is much higher than the performance of logistic regression models [16]. Similarly, deep regression models have been utilized in financial applications to predict a credit default risk with a high accuracy of 92.8%, greatly enhancing its risk assessment capacity [17]. In high-stakes environments, deep learning models can accurately predict risk scores by giving interpretable feature rankings that make them useful tools for deciding.

Despite the promising results, the study has not yet overcome some challenges in applying deep learning models to estimate risk scores. The first major concern is the risk of overfitting on small datasets. If sufficiently large and diverse datasets are available, deep learning models can be flexible and highly effective at capturing complex patterns [18]. The regularization techniques, cross-validation, and data augmentation can be used to mitigate this problem and improve model generalization.

Another challenge is the interpretability of deep learning models. While SHAP values and other techniques increase transparency, deep learning models remain a "black box" mostly because they can be extremely complex [19]. Future research should be on designing more interpretable AI frameworks that would have good predictive performance while at the same time guaranteeing trustworthy and explainable model outputs.

Moreover, those real-world applications also have a high computational cost. Transformer-based models are highly accurate and expensive regarding the data needed [20]. Thus, they cannot be deployed in environments with limited resources. Lightweight architectures, like efficient neural networks and knowledge distillation techniques, can be explored to improve these models for practical use.

Overall, this study confirms that regression-based deep learning models provide much improvement over traditional methods regarding risk score estimation and feature ranking. They are a useful tool across domains since they allow us to increase predictive accuracy and provide refined feature importance ranking. Nevertheless, taking advantage of these models in the real world will hinge on computational efficiency, interpretability, and dataset limitation. Future research can look into hybrid techniques that combine Deep Learning with domain knowledge to develop more robust and scalable risk assessment frameworks.

CONCLUSION

This study conducted a comprehensive meta-analysis of deep learning regression-based models for risk score estimation and feature ranking compared to traditional machine learning methods. Findings show deep learning models (particularly transformer-based architectures) significantly boost predictive predictive accuracy and feature interpretability. In particular, transformer-based models had an R^2 value of 0.95 and a minimal mean squared error of 0.007, outperforming standard methods such as linear regression and random forest for risk assessment.

Feature ranking analysis qualified deep learning models to be superior to others in identifying key risk factors and accurately identifying the top variables such as credit score, income level, and debt-to-income ratio. Unlike traditional methods like LASSO regression, transformer models enabled more refined and dynamic feature importance ranking than attention-based methods. This improvement improves model transparency and supports decision-making in the high-risk domain of finance and healthcare.

Deep learning models can provide both very high accuracy and feature ranking improvement, but at a cost—high computation costs and the possibility of overfitting. The practical deployment of these models in a resource-constrained environment is emphasized in the study and requires optimization techniques like model pruning and quantization. In addition, SHAP values and explainable AI frameworks that aid in model interpretation are important to help bring it to a wider audience.

This research confirms that regression-based deep learning models are a powerful framework for performing risk score estimation and feature ranking. Future work could make the computation more efficient, include more precise testing on real-world problems, and incorporate more domain knowledge to develop more robust, scalable, and – at least partially – interpretable risk assessment models. These further advances will bring better decision-making into fields that rely on accurate risk estimation.

Funding

This work was supported by Hainan Vocational College of Politics and Law College-level Education Reform Project Funding(YJJG2024-09)

REFERENCES

[1] Allied Market Research, "Predictive Analytics Market Size, Share, Competitive Landscape and Trend Analysis Report, by Component, by Enterprise Size, by Deployment Type, by Industry Vertical : Global Opportunity Analysis and Industry Forecast, 2024-2032," Allied Market Research, 2024. https://www.alliedmarketresearch.com/predictive-analytics-market

[2] D. Memic, "Assessing Credit Default using Logistic Regression and Multiple Discriminant Analysis:Empirical Evidence from Bosnia and Herzegovina," Interdisciplinary Description of Complex Systems, vol. 13, no. 1, pp. 128–153, 2015, doi: https://doi.org/10.7906/indecs.13.1.13.

[3] R. A. A. Mary and T. Ramaprabha, "Predicting Heart Disease Algorithm Using DNN and MNN in Deep Learning," International Research Journal on Advanced Engineering Hub (IRJAEH), vol. 2, no. 04, pp. 783–792, Apr. 2024, doi: https://doi.org/10.47392/irjaeh.2024.0110.

[4] A. Passemiers, P. Folco, D. Raimondi, G. Birolo, Y. Moreau, and Piero Fariselli, "A quantitative benchmark of neural network feature selection methods for detecting nonlinear signals," Scientific Reports, vol. 14, no. 1, Dec. 2024, doi: https://doi.org/10.1038/s41598-024-82583-5.

[5] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," Entropy, vol. 23, no. 1, p. 18, Dec. 2020, doi: https://doi.org/10.3390/e23010018.

[6] B. T. Johnson and E. A. Hennessy, "Systematic Reviews and meta-analyses in the Health sciences: Best Practice Methods for Research Syntheses," Social Science & Medicine, vol. 233, no. 233, pp. 237–251, Jul. 2019, doi: https://doi.org/10.1016/j.socscimed.2019.05.035.

[7] G. Ciulla and A. D'Amico, "Building energy performance forecasting: A multiple linear regression approach," Applied Energy, vol. 253, p. 113500, Nov. 2019, doi: https://doi.org/10.1016/j.apenergy.2019.113500.

[8] P. Shojaee, K. Meidani, B. F. Amir, and C. Reddy, "Transformer-based Planning for Symbolic Regression," Advances in Neural Information Processing Systems, vol. 36, pp. 45907–45919, Dec. 2023, Accessed: Mar. 17, 2025. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/hash/8ffb4e3118280a66b192b6f06e0e2596-Abstract-Conference.html

[9] X.-H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting," Water, vol. 11, no. 7, p. 1387, Jul. 2019, doi: https://doi.org/10.3390/w11071387.

[10] R. Caetano, J. M. Oliveira, and P. Ramos, "Transformer-Based Models for Probabilistic Time Series Forecasting with Explanatory Variables," Mathematics, vol. 13, no. 5, p. 814, Feb. 2025, doi: https://doi.org/10.3390/math13050814.

[11] Y. Meng, J. Jiang, J. Wu, and D. Wang, "Transformer-based deep learning models for predicting permeability of porous media," Advances in Water Resources, vol. 179, p. 104520, Aug. 2023, doi: https://doi.org/10.1016/j.advwatres.2023.104520.

[12] A. Kermani, H. Irani, and E. Zeraatkar, "Energy-Efficient Transformer Inference: Optimization Strategies for Time Series Classification," Arxiv.org, 2025. https://arxiv.org/html/2502.16627v1 (accessed Mar. 17, 2025).

[13] P. W. Tien, S. Wei, J. Darkwa, C. Wood, and J. K. Calautit, "Machine Learning and Deep Learning Methods for Enhancing Building Energy Efficiency and Indoor Environmental Quality – A Review," Energy and AI, vol. 10, p. 100198, Nov. 2022, doi: https://doi.org/10.1016/j.egyai.2022.100198.

[14] Shams Forruque Ahmed et al., "Deep learning modelling techniques: current progress, applications, advantages, and challenges," Artificial Intelligence Review, vol. 56, no. 1, Apr. 2023, doi: https://doi.org/10.1007/s10462-023-10466-8.

[15] H. Lee, J. Wang, and B. Leblon, "Using Linear Regression, Random Forests, and Support Vector Machine with Unmanned Aerial Vehicle Multispectral Images to Predict Canopy Nitrogen Weight in Corn," Remote Sensing, vol. 12, no. 13, p. 2071, Jun. 2020, doi: https://doi.org/10.3390/rs12132071.

[16] M. Tsai, K. Chen, and P. Chen, "Harnessing Electronic Health Records and Artificial Intelligence for Enhanced Cardiovascular Risk Prediction: A Comprehensive Review," Journal of the American Heart Association, Mar. 2025, doi: https://doi.org/10.1161/jaha.124.036946.

[17] V. Chang, S. Sivakulasingam, H. Wang, S. T. Wong, M. A. Ganatra, and J. Luo, "Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers," Risks, vol. 12, no. 11, p. 174, Nov. 2024, doi: https://doi.org/10.3390/risks12110174.

[18] I. D. Mienye and T. G. Swart, "A Comprehensive Review of Deep Learning: Architectures, Recent Advances, and Applications," Information, vol. 15, no. 12, p. 755, Nov. 2024, doi: https://doi.org/10.3390/info15120755.

[19] GeeksforGeeks, "Why Deep Learning is Black Box," GeeksforGeeks, Jul. 22, 2024. https://www.geeksforgeeks.org/why-deep-learning-is-black-box/

[20] L. Torbarina, T. Ferkovic, L. Roguski, V. Mihelcic, B. Sarlija, and Z. Kraljevic, "Challenges and opportunities of using transformer-based multi-task learning in NLP through ML lifecycle: A position paper," Natural language processing journal, pp. 100076–100076, Apr. 2024, doi: https://doi.org/10.1016/j.nlp.2024.100076.