Machine Learning Research
2025; 10(2): 151-159
http://www.sciencepublishinggroup.com/j/mlr
doi: 10.11648/j.mlr.20251002.18

A Deep Learning Approach to Polygenic Risk Prediction of Kidney Stone Formation

Amr Salem, Anirban Mondal*

Department of Mathematics, Applied Mathematics, and Statistics, Case Western Reserve University, Cleveland, USA

Email address:
axm912@case.edu (Anirban Mondal)

To cite this article:
Amr Salem, Anirban Mondal. (2025). A Deep Learning Approach to Polygenic Risk Prediction of Kidney Stone Formation. *Machine Learning Research, 10*(2), 151-159. https://doi.org/10.11648/j.mlr.20251002.18

Received: 22 November 2025 ; **Accepted:** 06 December 2025 ; **Published:** 19 December 2025

Abstract: Kidney stone disease affects millions of people worldwide, with both environmental and genetic factors contributing to individual susceptibility. While Genome-Wide Association Studies (GWAS) have successfully identified numerous single-nucleotide polymorphisms (SNPs) associated with kidney stone risk, translating these findings into accurate and clinically useful prediction tools remains a significant challenge. Polygenic Risk Scores (PRS) provide a framework for quantifying an individual's genetic predisposition, but traditional PRS approaches often fail to account for complex interactions among variants and the non-linear effects present in genomic data. In this study, we investigate the potential of deep learning techniques, specifically Convolutional Neural Networks (CNNs), to improve PRS models for predicting kidney stone susceptibility. Our approach integrates careful SNP selection, genotype filtering, and CNN-based modeling to address challenges such as data imbalance, genomic noise, and high-dimensional feature spaces. We benchmark our CNN-based framework against conventional machine learning models, including logistic regression, random forests, and support vector machines, demonstrating superior performance in terms of classification accuracy and ROC-AUC. These findings underscore the promise of deep learning-enhanced PRS models to provide more accurate genetic risk predictions for kidney stones. The research highlights the potential of integrating advanced computational approaches with genomics to advance precision medicine, improve patient stratification, and guide preventive strategies for at-risk individuals.

Keywords: Convolutional Neural Networks, Genome-Wide Association Studies, Polygenic Risk Score

1. Introduction

Kidney stones are a prevalent and burdensome health condition, affecting approximately one in ten individuals worldwide [8]. While environmental factors such as diet, hydration, and lifestyle strongly influence disease onset, genetic predisposition also plays a significant role in determining individual susceptibility [12]. Genome-Wide Association Studies (GWAS) have identified numerous single-nucleotide polymorphisms (SNPs) associated with kidney stone risk, yet translating these findings into clinically actionable prediction tools remains a major challenge [5].

Polygenic Risk Scores (PRS) provide a promising avenue for quantifying genetic susceptibility by aggregating the effects of many variants into a single predictive metric [3]. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), offer new opportunities to refine PRS models by capturing complex, non-linear relationships among genomic features [10]. Such models have demonstrated improved predictive performance across various complex diseases, including breast cancer, where deep neural networks have enhanced PRS estimation [4]. However, the application of deep learning-enhanced PRS frameworks to kidney stone prediction remains largely unexplored.

Despite their potential, deep learning methods face practical challenges when applied to large-scale genomic data, including class imbalance, noise, and the computational burden of processing high-dimensional inputs [7]. To address these challenges, we develop a CNN-based framework trained on a curated set of kidney stone-associated SNPs. Our approach integrates SNP selection, genotype filtering, and model optimization to investigate whether deep learning can

improve risk prediction in a PRS setting.

In this study, we analyze genotypic data derived from a recent GWAS on kidney stone susceptibility. To identify relevant genetic variants, we utilized GWAS summary statistics from Hao et al. (2023), which uncovered several novel loci associated with kidney stone formation [6]. These summary statistics informed the construction and weighting of our PRS model, ensuring close alignment with the most up-to-date genetic discoveries.

A CNN architecture was employed to model the non-linear relationships between the selected SNPs and kidney stone susceptibility. The model demonstrated consistent performance across validation and test sets, outperforming traditional machine learning approaches such as logistic regression, random forest, and support vector machines (SVM). Model implementation relied on TensorFlow for deep learning and Scikit-learn for classical machine learning techniques.

We further examined PRS distributions for both case and control groups. The case group exhibited a bimodal PRS distribution, identified using a Gaussian Mixture Model (GMM), suggesting the presence of genetically distinct subpopulations. In contrast, the control group displayed a unimodal distribution, indicative of a more homogeneous population. The bimodality observed in the case group highlights potential underlying genetic heterogeneity, warranting further investigation into subgroup-specific risk factors.

Overall, our findings underscore the potential of deep learning to enhance PRS-based risk prediction for kidney stones and contribute to the broader advancement of genomics-driven precision medicine. This study also establishes a foundation for future work aimed at integrating deep learning with large-scale genetic datasets to improve personalized risk assessment.

The remainder of the paper is organized as follows. Section 2 details the methodology, including data curation, preprocessing, and the CNN architecture. Section 3 presents model performance, comparisons with other methods, and interpretation of findings. Section 4 discusses the implications and limitations of the study. Finally, Section 5 provides concluding remarks and directions for future research.

2. Deep-Learning Methodology

This section details the methodology used in this study, including acquisition and preprocessing of GWAS data, construction of the polygenic risk score (PRS), linkage disequilibrium pruning, the Convolutional Neural Network (CNN) architecture, mathematical formulations, model training, evaluation procedures, and overall workflow.

2.1. Data Collection and Preprocessing

Genome-wide association study (GWAS) summary statistics for kidney stone susceptibility were obtained from Hao et al. (2023) [6]. These summary statistics included, for each single nucleotide polymorphism (SNP), the estimated effect size (β_j), corresponding standard error, p-value, allele frequency, and the specification of reference and effect alleles across millions of genomic loci. Such large-scale statistical information served as the foundation for SNP-level filtering, polygenic risk score (PRS) construction, and subsequent feature engineering procedures. The availability of high-resolution GWAS summary statistics enabled an informed, data-driven selection of genetic variants most likely to contribute to kidney stone susceptibility.

To construct individual-level polygenic risk scores, the conventional additive PRS framework was employed. For each participant i, the PRS was computed as

$$\mathrm{PRS}_i = \sum_{j=1}^{m} \beta_j \, G_{ij}, \tag{1}$$

where m denotes the number of SNPs retained for scoring, β_j is the GWAS-derived effect size associated with SNP j, and $G_{ij} \in \{0, 1, 2\}$ represents the genotype dosage corresponding to the count of effect alleles carried by individual i. This formulation assumes additive genetic effects and allows for a compact representation of an individual's cumulative genetic predisposition. The PRS not only served as an interpretable aggregate metric but also aided in validating the consistency of SNP filtering and downstream modeling pipelines.

To mitigate redundancy arising from correlated genomic markers, linkage disequilibrium (LD) pruning was performed on the full SNP dataset. Pairwise LD was quantified using the squared correlation coefficient,

$$r^2_{jk} = \frac{\mathrm{Cov}(G_j, G_k)^2}{\mathrm{Var}(G_j)\,\mathrm{Var}(G_k)}, \tag{2}$$

and SNP pairs exhibiting values exceeding the threshold $r^2 > 0.8$ were considered highly correlated. In such cases, one SNP from each LD block was removed to ensure that only largely independent variants were retained. This pruning step effectively reduced multicollinearity within the feature set and prevented inflation of model weights due to redundant predictors.

The individual-level genotype dataset used for model training and evaluation was obtained from the All of Us Research Hub [1], consisting of 560 participants with high-quality SNP array data. All individuals self-identified as White, and cases and controls were explicitly age- and gender-matched to minimize potential demographic confounding. Cases were identified based on the presence of the ICD-10 diagnosis code N20 (kidney and ureteral calculi), whereas controls were randomly selected from participants without any recorded N20 diagnosis. After applying standard quality-control filters-including call-rate thresholds and checks for allele frequency consistency-the curated dataset was partitioned into two subsets. A total of 500 individuals were used exclusively for model training and validation, while 60 individuals were reserved as a held-out test set.

2.2. CNN Model Architecture

The convolutional neural network (CNN) developed for this study was designed to model the complex, non-linear dependencies among single nucleotide polymorphisms (SNPs) contributing to kidney stone susceptibility. CNNs have proven effective in genomics for identifying local patterns such as regulatory motifs, chromatin accessibility signals, and transcription factor binding sites [2, 10, 11, 13]. These characteristics motivate their use for polygenic risk prediction, where short-range genomic interactions and haplotype-like structures may be relevant.

In this study, each individual's genotype vector was represented as a one-dimensional sequence and reshaped into an input tensor of shape $(L, 1)$, where L denotes the number of filtered SNPs. Convolutional filters were employed to scan across adjacent SNP positions, enabling the model to automatically detect short-range dependencies such as local LD structure or variant co-occurrence patterns. Max-pooling layers reduced feature-map dimensionality, providing translational robustness and improved computational efficiency. Dropout layers were included to mitigate overfitting, which is of particular importance given the modest sample size.

The exact CNN implementation used in this study consisted of the following sequence of layers: a *Conv1D* layer with 64 filters and kernel size 3 (ReLU activation), followed by *MaxPooling1D* (pool size = 2) and a *Dropout* layer (rate = 0.2). A second *Conv1D* layer with 32 filters and kernel size 3 (ReLU activation) was applied next, followed by another *MaxPooling1D* (pool size = 2) and *Dropout* layer (rate = 0.2). The resulting feature maps were flattened and passed into a *Dense* layer with 32 units (ReLU activation), and finally into a *sigmoid-activated output layer* with 1 unit for binary risk prediction.

Mathematically, each convolutional layer computes:

$$h_k^{(l)}(t) = f\left(\sum_{i=0}^{K-1} w_{k,i}^{(l)}\, x_{t+i}^{(l-1)} + b_k^{(l)}\right), \tag{3}$$

where K is the kernel size, $w_{k,i}^{(l)}$ and $b_k^{(l)}$ are learnable parameters, and $f(\cdot)$ denotes the ReLU activation function. By sliding these filters across the SNP sequence, the model captures position-dependent local interactions that may not be easily modeled by traditional machine-learning approaches.

The model was trained using the Adam optimizer (learning rate = 0.001), binary cross-entropy loss, batch size = 8, and up to 250 epochs with early stopping (patience = 10). These hyperparameters reflect the exact configuration used in the accompanying TensorFlow implementation and were selected to balance computational efficiency with effective convergence.

Overall, this architecture leverages CNNs' strengths in hierarchical feature extraction, parameter efficiency through weight sharing, and the ability to capture subtle local genomic patterns. These characteristics make CNNs a compelling methodological choice for SNP-based prediction tasks such as kidney stone susceptibility.

2.3. Training and Validation

The convolutional neural network (CNN) was trained using the Adam optimizer with a learning rate of 0.001, which provides adaptive moment-based updates to network parameters and has been shown to converge efficiently across a wide range of deep learning tasks [9]. Binary cross-entropy was employed as the loss function to quantify the discrepancy between predicted probabilities and true binary labels, as it is well suited for dichotomous classification problems. To prevent overfitting and to ensure that training was halted once generalization performance ceased to improve, early stopping was implemented based on the validation loss. Model generalizability was further assessed through five-fold cross-validation, wherein the training set was partitioned into five subsets, each serving as a validation set in turn while the remaining subsets were used for training. This procedure allowed for robust estimation of model performance across different subsets of the data and mitigated potential biases arising from specific sample partitions. To prevent data leakage, the 60-sample test set was held out prior to any cross-validation or hyperparameter tuning. Five-fold cross-validation was performed strictly within the 500-sample training set. For each fold, the model was trained on 80% of the fold and validated on the remaining 20%. Feature scaling and preprocessing steps, when applied, were fit only on the training portion of each fold and then applied to the validation portion to avoid information leakage. No individual appeared in more than one role (train/validation/test) during evaluation.

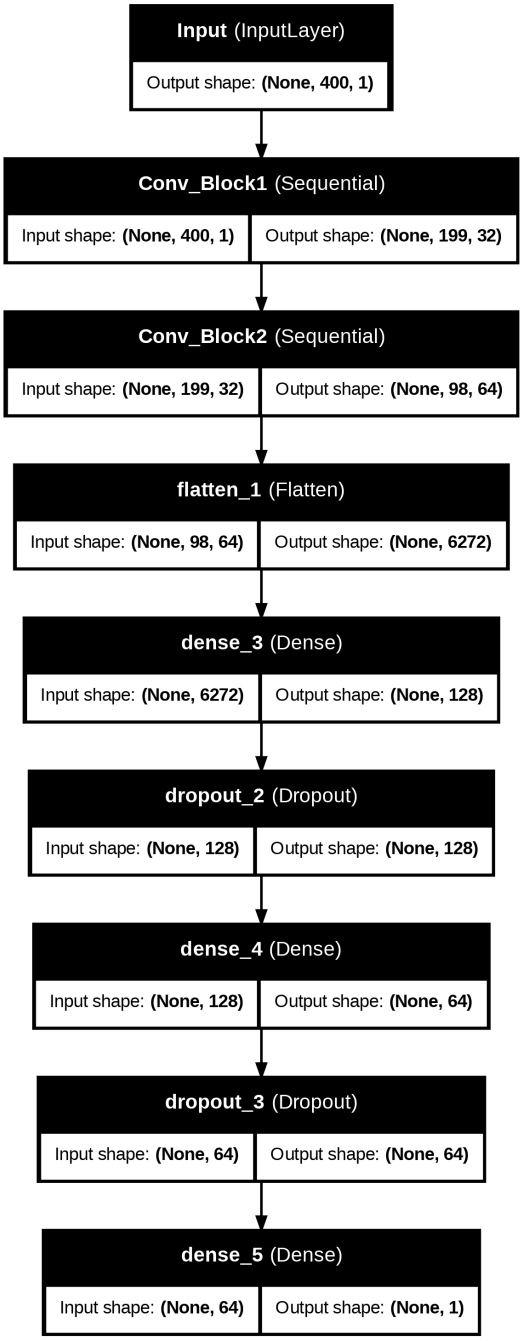
2.4. Model Comparison

To evaluate the predictive advantage of the CNN, we benchmarked its performance against several classical machine learning approaches commonly used in genomic prediction studies. Logistic regression was included as a linear baseline model and was regularized with an L2 penalty to reduce overfitting and stabilize coefficient estimates. Random forests with 100 trees were used to capture non-linear interactions among SNP features, while support vector machines (SVM) with a radial basis function (RBF) kernel modeled complex decision boundaries. Gradient boosting, implemented via the XGBoost framework with a learning rate of 0.1 and a maximum tree depth of three, served as an additional non-linear baseline. Evaluation was conducted exclusively using the area under the ROC curve (ROC-AUC), which was the predefined primary metric for all models.

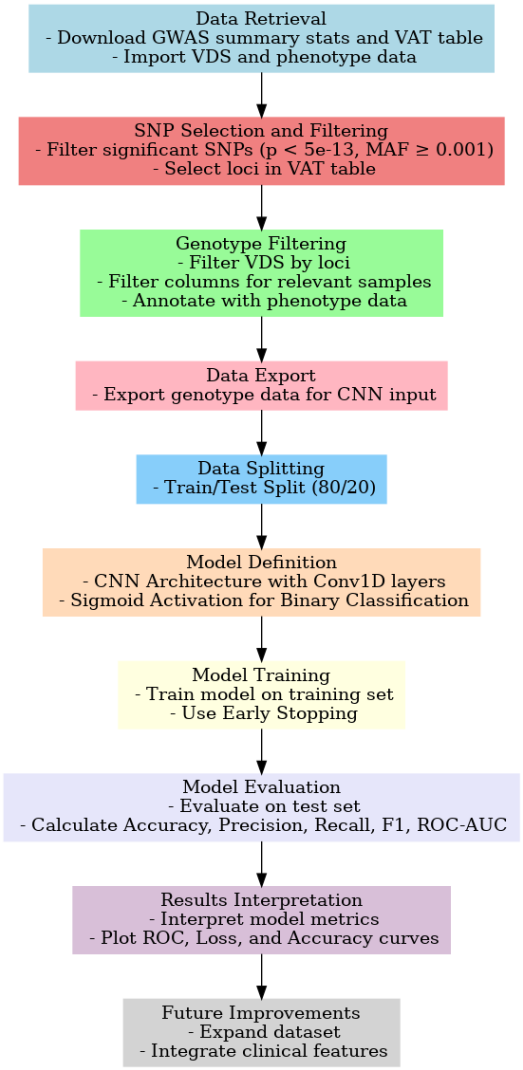
2.5. Methodology Workflow

The overall workflow of this study is summarized in Figure 1b and reflects an integrated pipeline for polygenic risk prediction using deep learning. Initially, GWAS summary statistics were acquired and harmonized to ensure consistency in reference alleles, effect sizes, and variant identifiers across datasets. SNPs were then selected based on statistical

significance thresholds and pruned to remove highly correlated variants using linkage disequilibrium measures, thereby reducing redundancy and multicollinearity. Individual-level genotype and phenotype data were preprocessed to handle missing values, encode genotypes numerically, and standardize input features. Polygenic risk scores were computed and arranged into matrices suitable for CNN input, preserving the sequential structure of SNPs. The CNN model was trained and validated following the procedure described above, and its predictive performance was rigorously evaluated using multiple metrics. Comparisons were then made against baseline classical machine learning models to benchmark the added value of deep learning in capturing complex, non-linear genetic interactions. Finally, model outputs were interpreted to explore feature contributions and potential avenues for future methodological improvements, including the integration of additional omics data or alternative neural architectures.



(a) *Schematic illustration of the CNN architecture.*



(b) *Workflow Overview. Step-by-step illustration of SNP selection, genotype filtering, CNN model training, evaluation, and interpretation.*

Figure 1. *The CNN architecture and the workflow of the methodology.*

3. Results

To identify the most predictive SNPs for kidney stone susceptibility, GWAS summary statistics were filtered using varying p-value thresholds. These thresholds were evaluated empirically across multiple model runs, and the threshold of 10^{-8} consistently produced the strongest validation and test performance. Table 1 summarizes accuracy across all tested thresholds, demonstrating stable behavior of the model. Given its superior performance, the 10^{-8} threshold was selected for the final model and is the result reported here.

Table 1. *Validation and test accuracy for models trained with different GWAS p-value thresholds. The model using a threshold of 10^{-8} achieved the highest performance on both validation and test sets and was selected for subsequent analyses.*

P-value threshold	Validation Accuracy	Test Accuracy
10^{-8}	0.62	0.62
10^{-9}	0.55	0.57
10^{-10}	0.575	0.58
10^{-11}	0.60	0.54
10^{-12}	0.525	0.58

The selected model achieved a validation accuracy of 62% and a test accuracy of 61.7%, indicating that the model generalizes reasonably well to unseen data.

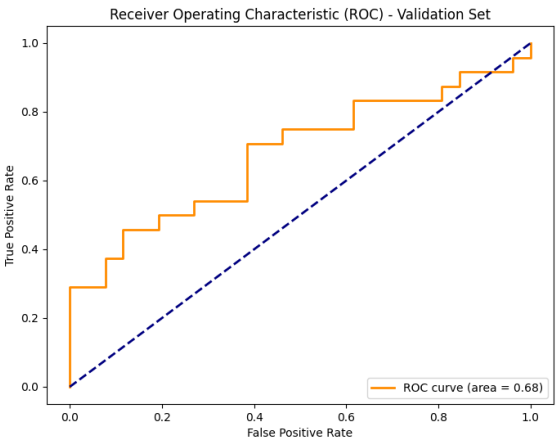
3.1. CNN Model Performance

The convolutional neural network (CNN) was trained using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy loss, with early stopping based on validation loss to prevent overfitting. Five-fold cross-validation was employed to evaluate model robustness. The CNN achieved a validation accuracy of 62% with a ROC-AUC of 0.68, a precision of 0.60, a recall of 0.63, and an F1-score of 0.61. On the held-out test set, the model achieved comparable performance, with an accuracy of 61.7%, ROC-AUC of 0.68, precision of 0.63, recall of 0.57, and an F1-score of 0.60, reflecting modest overfitting but stable precision across datasets. Loss values were 0.72 and 0.74 for validation and test sets, respectively. These results indicate that the CNN can reliably capture complex, non-linear interactions among SNPs that contribute to kidney stone risk.

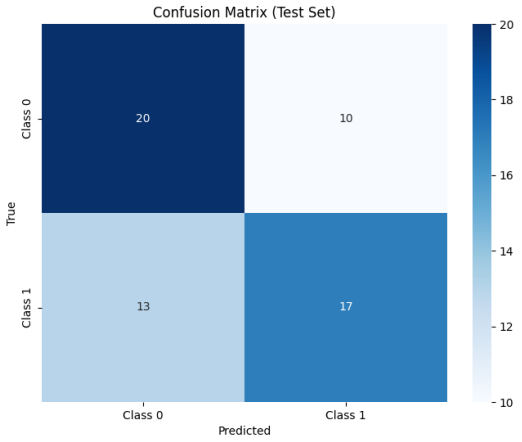
Figure 2 presents the ROC curve for the validation set, the confusion matrix for the test set, and the PRS distributions for cases and controls. The ROC curve illustrates the CNN's ability to discriminate high- and low-risk individuals, while the confusion matrix demonstrates the balance between sensitivity and specificity.

3.2. Polygenic Risk Score Distribution

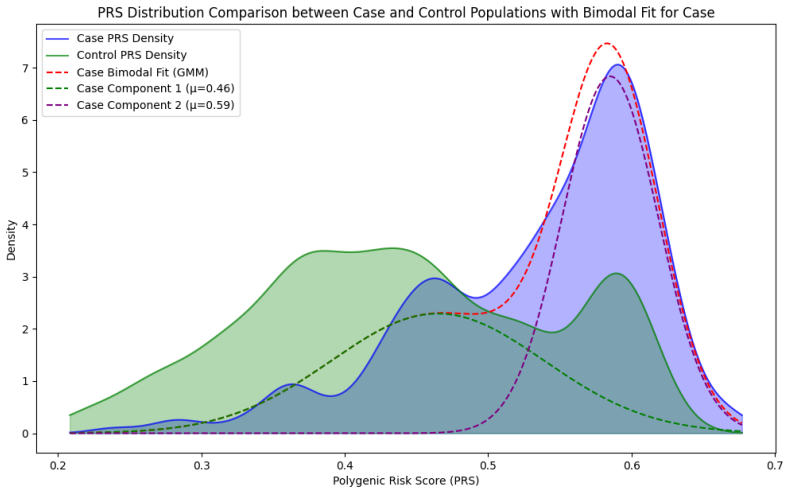
The distribution of polygenic risk scores was further analyzed to explore the genetic architecture of kidney stone susceptibility. Kernel Density Estimation (KDE) revealed a visible shift between cases and controls, although some overlap persisted. Within the case group, Gaussian Mixture Model (GMM) fitting suggested a bimodal pattern, indicating the possibility that certain individuals may carry distinct combinations of higher-risk alleles. In contrast, the control group displayed a unimodal distribution, reflecting a more homogeneous genetic background. These observations imply that PRS may help highlight heterogeneity in genetic risk among cases. We note, however, that this apparent separation and the suggested bimodal pattern are observational and should be interpreted cautiously; they may indicate potential genetic subgroups, but confirming such structure would require validation in larger and more diverse cohorts.



(a) *ROC curve for the validation set. AUC = 0.68.*



(b) *Confusion matrix for the test set. Precision = 63%, Recall = 57%.*



(c) *Distribution of polygenic risk scores for cases and controls using kernel density estimation. The case group exhibits a bimodal distribution.*

Figure 2. *Evaluation of CNN model performance and polygenic risk score (PRS) distributions. Panel (a) shows the ROC curve for the validation set, panel (b) displays the confusion matrix for the test set, and panel (c) illustrates PRS distributions for cases and controls, highlighting a bimodal pattern in the case group.*

3.3. Comparative Analysis with other Machine Learning Models

To benchmark the predictive advantage of the CNN, performance was compared with logistic regression, random forests, support vector machines (SVM), and gradient boosting (XGBoost). Logistic regression achieved an AUC of 0.48, random forests 0.62, SVM 0.64, and gradient boosting 0.61. The CNN outperformed all classical models with an AUC of 0.68, emphasizing its ability to capture complex, non-linear SNP interactions that are not easily modeled by linear or tree-based methods. Figure 3 presents a comparative view of the ROC curves and highlights the superior discriminative ability of the CNN.

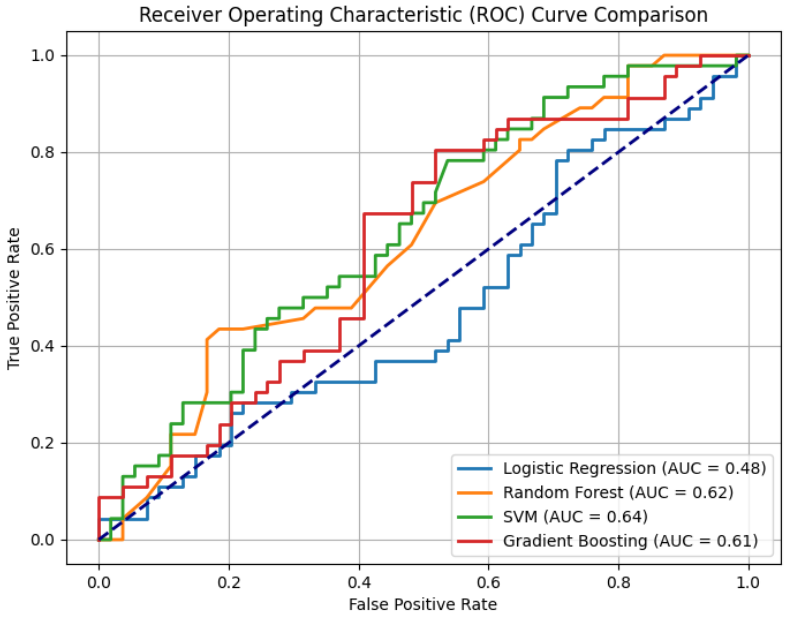


Figure 3. *Comparison of ROC curves between the CNN and classical machine learning models. The CNN achieved the highest AUC (0.68), outperforming support vector machines (0.64), random forests (0.62), gradient boosting (0.61), and logistic regression (0.48), demonstrating its capability to model complex, non-linear SNP interactions.*

4. Discussion

This study investigated the use of Convolutional Neural Networks (CNNs) for predicting kidney stone susceptibility from genomic data. By integrating polygenic risk scores (PRS) derived from genome-wide association studies (GWAS), the model captured complex, non-linear interactions among genetic variants, highlighting the potential of deep learning for improving risk stratification in complex health outcomes. The CNN architecture, detailed in Section 2, combined convolutional and pooling layers to extract hierarchical genomic features, followed by fully connected layers for binary classification.

4.1. Model Performance and Evaluation

The CNN model demonstrated superior performance compared to traditional machine learning approaches, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and Gradient Boosting (see Section 3.3). These findings suggest that deep learning methods can effectively exploit high-dimensional genomic data, identifying intricate patterns that conventional models may overlook. Regularization strategies, such as dropout and early stopping, were instrumental in controlling overfitting, especially given the moderate sample size, and contributed to the robustness and stability of the model. The ROC-AUC analysis further confirmed that the CNN could reliably distinguish higher-risk individuals, validating the effectiveness of the learned representations.

4.2. PRS Distribution and Its Implications

Analysis of PRS distributions, derived from GWAS summary statistics as described in Section 2.1, provided additional insights into genetic susceptibility. The case group exhibited a bimodal PRS distribution, suggesting the presence of genetically distinct subpopulations with varying levels of risk. In contrast, the control group showed a unimodal distribution, reflecting greater homogeneity among individuals without kidney stone susceptibility. The partial overlap between case and control PRS distributions illustrates the inherent challenge of predicting complex traits and emphasizes the need for models that capture subtle genetic interactions. These findings also support the network's performance as measured by ROC-AUC, showing that the CNN successfully learned meaningful genomic patterns for risk prediction.

4.3. Comparison with Previous Studies

Prior research has primarily focused on individual SNPs associated with kidney stone risk [5, 12], while traditional PRS models combine multiple variants to improve risk estimation [3]. Our approach extends these frameworks by leveraging a comprehensive set of GWAS-derived SNPs [6] and employing CNNs to model non-linear, higher-order interactions among variants (see Section 2.2). This methodological advancement enables more nuanced risk prediction and demonstrates the added value of deep learning in genomics-driven precision medicine.

4.4. Limitations of the Study

Several limitations of this study should be acknowledged. First, the relatively small sample size of 500 individuals may limit the generalizability of the findings and increase the risk of overfitting, as the model could capture patterns specific to the training data rather than broader population trends. Second, the study cohort predominantly represents a single population, which raises concerns about the transferability of the model to ethnically and genetically diverse groups. Kidney stone susceptibility is influenced by a combination of genetic, environmental, and lifestyle factors, which can vary substantially across populations. Accordingly, unmeasured confounding factors, such as population stratification, gene-environment interactions, and lifestyle-related exposures, may influence the observed associations and cannot be fully accounted for in the current framework. As a result, predictive models trained on a single cohort may not achieve comparable performance in other demographic or ethnic groups. Furthermore, future studies leveraging whole-genome sequencing (WGS) would enable a more comprehensive assessment of the genetic architecture underlying kidney stone susceptibility. Future research incorporating larger, more heterogeneous datasets will be critical to validate these results, improve the robustness of the model, and enhance its applicability across diverse populations.

4.5. Implications for Genomic Risk Prediction

Overall, this study demonstrates that CNN-based models, when combined with PRS derived from comprehensive GWAS data and processed according to the workflow shown in Figure 1b, can provide meaningful insights into the genetic architecture of kidney stone susceptibility. The findings suggest that deep learning approaches complement traditional risk prediction methods, capturing complex interactions among genetic variants and informing more personalized strategies for prevention and early intervention.

4.6. Future Directions

The findings of this study open several promising avenues for future research. Expanding the dataset to include a more diverse population will be essential for improving model generalization and ensuring applicability across different ethnic and demographic groups. Collaborations with other research initiatives could facilitate access to larger and more heterogeneous datasets, enhancing the robustness of the predictions.

Incorporating additional data types, such as clinical measurements (e.g., blood markers, urinary composition) and environmental or lifestyle factors (e.g., diet, hydration), could further improve the predictive performance of the model. Integrating genetic, phenotypic, and lifestyle information

would enable a more comprehensive and individualized approach to kidney stone risk assessment.

Exploring alternative deep learning architectures, including recurrent neural networks (RNNs) or transformers, may enhance the model's capacity to capture complex patterns in genomic data. Further optimization through hyperparameter tuning, architectural refinement, and ensemble learning techniques could also improve prediction accuracy and reliability.

Ultimately, the long-term goal is to develop a clinically applicable tool for individualized kidney stone risk prediction. Such a tool could inform preventative strategies, guide personalized treatment plans, and reduce the overall burden of kidney stones in the population.

4.7. Clinical Applicability

While the proposed CNN-based PRS framework shows promise, several challenges hinder immediate clinical deployment. These include computational requirements for model training, the cost and availability of SNP-array genotyping, integration with electronic health record systems, model calibration for clinical decision-making thresholds, and the need for validation in multi-ethnic cohorts. As such, the present work should be viewed as methodological rather than directly translational.

5. Conclusion

This study highlights the utility of deep learning for genomic-based risk prediction, focusing on kidney stone susceptibility. By employing a Convolutional Neural Network to model complex interactions among SNPs from GWAS data, the proposed approach demonstrated superior predictive performance compared to traditional machine learning methods. Although the model achieved encouraging results, further work is necessary to enhance generalizability, mitigate dataset imbalances, and incorporate additional genetic, clinical, and environmental factors. Taken together, the results of this study should be interpreted as hypothesis-generating rather than conclusive. The modest sample size, demographic homogeneity, and exploratory nature of the modeling framework indicate that the observed performance patterns and PRS stratification signals require replication in larger and more diverse cohorts. Accordingly, the findings presented here are best viewed as preliminary evidence that motivates future, more comprehensive investigations into the genetic architecture of kidney stone susceptibility and the broader potential of deep learning for polygenic risk prediction.

ORCID

0009-0008-7161-7346 (Amr Salem)
0000-0002-4100-2366 (Anirban Mondal)

Abbreviations

GWAS	Genome-Wide Association Studies
SNP	Single-nucleotide Polymorphism
PRS	Polygenic Risk Scores
CNN	Convolutional Neural Network
SVM	Support Vector Machine

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] All of us research program. https://www.researchallofus.org/, 2025. Accessed: 2025-11-21.

[2] Babak Alipanahi, Andrew Delong, Matthew T Weirauch, and Brendan J Frey. Predicting the sequence specificities of dna-and rna-binding proteins by deep learning. Nature Biotechnology, 33(8): 831-838, 2015.

[3] J. Allen and K. Thompson. Polygenic risk scores: A comprehensive review. Genetic Epidemiology, 41: 169-182, 2017. https://doi.org/10.1002/gepi.22045

[4] A. Badré et al. Improving polygenic risk prediction using deep neural networks: Applicationto breast cancer. PLoS Genetics, 17(4): e1009522, 2021.

[5] E. Brown and F. Green. Genome-wide association studies on kidney stone susceptibility. Human Genetics, 137: 789-800, 2018. https://doi.org/10.1007/s00439-018-1923-7

[6] X. Hao, Z. Shao, N. Zhang, et al. Integrative genome-wide analyses identify novel loci associated with kidney stones and provide insights into its genetic architecture. Nature Communications, 14: 7498, 2023. https://doi.org/10.1038/s41467-023-43400-1

[7] H. He and E. Garcia. Learning from imbalanced data: Techniques and challenges. IEEE Transactions on Knowledge and Data Engineering, 21: 1263-1284, 2009. https://doi.org/10.1109/TKDE.2008.239

[8] A. Johnson and B. Smith. Global prevalence and risk factors for kidney stones. Journal of Nephrology, 35: 102-110, 2020. https://doi.org/10.1000/j.jneph.2020.01.005

[9] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. International Conference on Learning Representations, 2015.

[10] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. Nature, 521:436-444, 2015. https://doi.org/10.1038/nature14539

[11] Ryan et al. Poplin. A universal snp and small-indel variant caller using deep neural networks. Nature Biotechnology, 36(10): 983-987, 2018. 16

[12] C. Smith and D. Lee. Genetic insights into kidney stone formation. Nature Genetics, 47: 456-462, 2015. https://doi.org/10.1038/ng.3250

[13] Jian Zhou and Olga G Troyanskaya. Predicting effects of noncoding variants with deep learning-based sequence model. Nature Methods, 12(10): 931-934, 2015.