Metric-Enhanced Hybrid Kernel Probabilistic Neural Networks for Robust Classification

Abhishek Varshney[a]
Department of Mathematics
Indian Institute of Technology Indore
phd2401141003@iiti.ac.in

Mushir Akhtar[a]
Department of Mathematics
Indian Institute of Technology Indore
phd2101241004@iiti.ac.in

Mohd. Arshad[*]
Department of Mathematics
Indian Institute of Technology Indore
arshad@iiti.ac.in

M. Tanveer
Department of Mathematics
Indian Institute of Technology Indore
mtanveer@iiti.ac.in

***Abstract*—Probabilistic Neural Networks (PNNs) offer an interpretable and uncertainty-aware framework for classification by explicitly modeling class-conditional probability densities. However, classical PNNs and their existing variants rely on isotropic Euclidean distance and single-kernel Gaussian density estimation, which limits their ability to capture feature relevance, inter-feature dependencies, and heterogeneous data geometries commonly observed in real-world and imbalanced datasets. To address these limitations, this paper proposes a Metric-Enhanced Hybrid Kernel Probabilistic Neural Network (MEHK-PNN), which augments the classical PNN architecture through two complementary mechanisms. First, a metric-enhanced feature transformation reshapes the input space by encoding feature relevance and correlations, yielding a more discriminative similarity geometry for density estimation. Second, a hybrid-kernel Parzen estimator is introduced, combining a Gaussian kernel with a complementary similarity function to jointly capture local neighborhood structure and broader relational patterns. Two instantiations of the proposed framework are developed, incorporating cosine and sigmoid kernels as complementary components. Extensive experiments conducted on 35 benchmark KEEL datasets demonstrate that the proposed MEHK-PNN variants consistently outperform classical PNNs, recent probabilistic extensions, and strong discriminative baselines in terms of F1-score, stability, and statistical significance. Overall, the proposed framework preserves the probabilistic transparency and non-iterative training advantages of PNNs while significantly enhancing robustness and classification effectiveness under imbalanced and heterogeneous data conditions. Supplement file is available at https://github.com/anonymousabhivarsh/Supplement_MEHK_PNN.**

***Index Terms*—Probabilistic neural networks, metric learning, hybrid kernel, Parzen density estimation, robust classification.**

I. Introduction

MACHINE learning models have achieved remarkable success across a wide range of application domains [1, 2], including pattern recognition [3, 4], biomedical diagnosis [5, 6], and computer vision [7, 8]. Consequently, discriminative classifiers such as support vector machines (SVMs) [9, 10], random forests (RF) [11], gradient boosting methods including XGBoost [12], and multilayer perceptrons (MLPs) [13] have become standard baselines in practical classification tasks. Despite their strong empirical performance, these models are primarily optimized for point prediction accuracy and typically lack an explicit probabilistic formulation. As a result, they offer limited insight into predictive uncertainty, which is a critical requirement in high-stakes applications. Furthermore, the black-box nature of many widely adopted discriminative models often hampers interpretability and reduces trust in their predictions [14].

In contrast to purely discriminative approaches, probabilistic classifiers explicitly model data uncertainty and class-conditional distributions, thereby providing a principled foundation for decision making. Among such models, the Probabilistic Neural Network (PNN), introduced by Specht [15], is particularly appealing due to its transparent probabilistic formulation and analytically interpretable architecture. PNN estimates class-conditional probability density functions using Parzen window techniques and performs classification via Bayesian decision theory. This formulation naturally yields posterior probabilities and enables uncertainty-aware predictions, making PNN especially suitable for applications where reliability and interpretability are essential [16]. An additional advantage of PNN lies in its effectiveness under class-imbalanced learning scenarios. By explicitly estimating class-conditional densities, PNN does not rely solely on hard decision boundaries and is therefore less biased toward majority classes compared to many discriminative classifiers. The network architecture, comprising input, pattern, summation, and decision layers, is simple and analytically tractable, with each component admitting a clear statistical interpretation. These properties make PNN an attractive and robust alternative to more complex black-box learning models, particularly in imbalanced and uncertainty-sensitive classification tasks.

Over the past decade, several extensions of the classical PNN have been proposed to alleviate its limitations in probability density estimation and class discrimination. Competitive PNN (C-PNN) [17] introduces a competition mechanism to select representative kernels for each class, thereby reducing redundancy in density modeling and improving performance

[*]Corresponding author. [a]Equal contribution.

on multi-clustered data. Weighted PNN (W-PNN) [18] incorporates adaptive weighting factors between the pattern and summation layers to emphasize informative samples while suppressing noisy contributions. To explicitly address class imbalance, Improved PNN (I-PNN) [19] modifies the density aggregation process to account for unequal class frequencies, leading to more reliable posterior probability estimates. More recently, Skew-N-PNN [20] replaces the conventional Gaussian kernel with a skew-normal distribution to enable asymmetric density modeling and enhance minority-class recognition under skewed data distributions.

Although prior studies have explored various extensions of PNNs for imbalanced and asymmetric learning, they largely emphasize changes to aggregation strategies or kernel shapes, leaving the underlying similarity modeling largely unchanged. In particular, most PNN variants continue to rely on a single-kernel formulation combined with simple similarity measures, which fundamentally limits their capacity to capture complex feature dependencies and heterogeneous data geometries.

In conventional PNN formulations, similarity between samples is computed directly in the original feature space using Euclidean distance within a Gaussian kernel. This design implicitly assumes that all features contribute equally to the classification task, inter-feature correlations are negligible, and class-conditional distributions exhibit approximately isotropic structure. Such assumptions are often violated in real-world datasets, which typically exhibit heterogeneous feature relevance, strong statistical dependencies among variables, heavy-tailed or skewed distributions, and diverse local geometries across different classes. Under these conditions, Euclidean distance becomes an inadequate measure of similarity, leading to inaccurate density estimation, degraded class discrimination, and increased sensitivity to noise and irrelevant attributes. Furthermore, restricting the Parzen estimator to a single Gaussian kernel significantly limits its expressive power, as it primarily captures local smoothness while failing to represent broader structural relationships and global similarity patterns present in complex data. These limitations collectively hinder the robustness and generalization ability of classical and extended PNN models, particularly in imbalanced learning settings.

To address the aforementioned limitations while preserving the simplicity and probabilistic transparency of classical PNNs, this paper proposes an enhanced framework termed the Metric-Enhanced Hybrid Kernel Probabilistic Neural Network (MEHK-PNN). The proposed approach augments the original PNN architecture through two complementary mechanisms: a metric-enhanced feature transformation and a hybrid-kernel Parzen density estimation layer. The metric-enhanced transformation learns a discriminative similarity space by explicitly modeling feature relevance and inter-feature correlations, thereby reshaping the input geometry and alleviating the restrictive isotropic assumptions imposed by Euclidean distance. In parallel, the hybrid kernel formulation enriches the Parzen density estimator by combining complementary kernel functions within a unified similarity framework, allowing the model to simultaneously capture local neighborhood structure and broader relational patterns in the data. This hybrid similarity modeling enhances robustness to noise, improves discrimination under overlapping class distributions, and strengthens minority-class density estimation in imbalanced learning scenarios.

The main contributions of this work are summarized as follows:

1) We propose a metric-enhanced hybrid-kernel probabilistic neural network (MEHK-PNN) that simultaneously addresses geometric inadequacy and limited similarity expressiveness in classical PNNs.
2) We introduce a metric-based feature transformation that constructs a discriminative similarity space tailored for probabilistic density estimation under feature correlation and heterogeneity.
3) We employ a hybrid kernel Parzen estimator that improves robustness and minority-class density modeling in imbalanced classification settings.
4) We conduct extensive experimental evaluations on 35 KEEL benchmark datasets, reporting mean F1-score along with standard deviation, and further validate the results using ranking-based statistical significance analysis via the Friedman test followed by Nemenyi post-hoc comparisons.

The remainder of this paper is organized as follows. Section II introduces the notation and reviews the classical PNN framework. Section III presents the proposed Metric-Enhanced Hybrid Kernel Probabilistic Neural Network (MEHK-PNN) in detail. Experimental results are discussed in Section IV. Finally, Section V concludes the paper.

II. Related Work

In this section, we introduce the notations and review the classical PNN framework.

A. Notations

Throughout this paper, scalars are denoted by lowercase letters (e.g., x), vectors by bold lowercase letters (e.g., $\mathbf{x}$), and matrices by bold uppercase letters (e.g., $\mathbf{L}$). The training dataset is represented as $\mathcal{D} = \{(\mathbf{x}_k, y_k)\}_{k=1}^{N}$, where $\mathbf{x}_k \in \mathbb{R}^d$ denotes the input feature vector and $y_k \in \{1, 2, \ldots, C\}$ is the corresponding class label. The function $K(\cdot, \cdot)$ represents a generic kernel used to measure similarity between samples, and $\|\cdot\|$ denotes the Euclidean norm.

B. Probabilistic Neural Network (PNN)

PNN [15] is a feedforward classification model grounded in Bayesian inference and nonparametric density estimation. Unlike gradient-based neural networks, PNN estimates class-conditional probability density functions directly from training data using kernel-based Parzen window techniques and performs classification according to the Bayesian decision rule. This formulation yields probabilistic outputs that quantify prediction confidence and provides an interpretable decision mechanism.

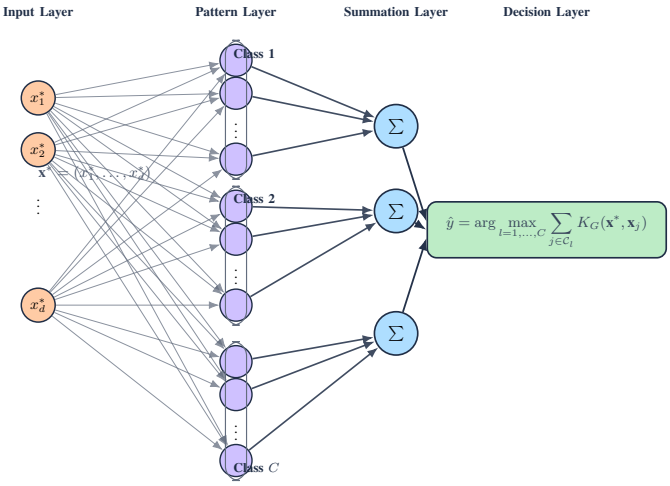

Fig. 1: Four-layer PNN architecture: similarities are aggregated per class and the label is predicted by maximum class density.

A classical PNN consists of four layers: input, pattern, summation, and decision, as illustrated in Fig. 1. Given a query sample $\mathbf{x}^* \in \mathbb{R}^d$, the input layer forwards it to the pattern layer, where each neuron corresponds to a training sample $\mathbf{x}_k$. A similarity score between a query sample $\mathbf{x}^*$ and a training sample x_k is computed using a Gaussian kernel,

$$K_G(\mathbf{x}^*, \mathbf{x}_k) = \exp\left(-\frac{\|\mathbf{x}^* - \mathbf{x}_k\|^2}{2\sigma^2}\right), \tag{1}$$

where σ controls the smoothness of the density estimate. For simplicity, the normalization constant is omitted, as it does not affect the classification decision in PNN due to cancellation in posterior probability computation.

The summation layer aggregates kernel responses within each class to approximate the class-conditional density. For class j, this estimate is given by

$$f_j(\mathbf{x}^*) = \frac{1}{n_j} \sum_{\mathbf{x}_k \in C_j} K_G(\mathbf{x}^*, \mathbf{x}_k), \tag{2}$$

where n_j denotes the number of training samples in class j and C_j is the corresponding index set.

Finally, the decision layer computes posterior class probabilities using Bayes' theorem,

$$P(y = j \mid \mathbf{x}^*) = \frac{P(y = j)\, f_j(\mathbf{x}^*)}{\sum_{\ell=1}^{C} P(y = \ell)\, f_\ell(\mathbf{x}^*)}, \tag{3}$$

and assigns the class label with the maximum posterior probability. In many applications, uniform class priors are assumed, such that classification is governed solely by the estimated class-conditional densities.

III. Proposed Method

This section presents the proposed Metric-Enhanced Hybrid Kernel Probabilistic Neural Network (MEHK-PNN), which extends the classical PNN by improving similarity modeling and probabilistic density estimation. The proposed framework enhances the original architecture through two complementary components: (i) a metric-enhanced feature transformation that reshapes the input geometry for improved distance computation, and (ii) a hybrid-kernel Parzen density estimator that increases representational flexibility while preserving the nonparametric and probabilistic nature of PNNs.

A. Metric-Enhanced Feature Transformation

In classical PNNs, similarity between samples is computed directly in the original feature space using Euclidean distance, implicitly assuming equal feature relevance, weak inter-feature dependence, and isotropic class distributions. However, these assumptions are often violated in real-world datasets, where heterogeneous feature importance and correlated structures can degrade the reliability of distance-based density estimation.

To address this limitation, we adopt a more intuitive approach: instead of computing similarity in the original feature space, we first reshape the space so that important features are emphasized and correlated features are properly aligned. This transformation enables distance computations to better capture the intrinsic structure of the data, leading to more reliable similarity estimation. MEHK-PNN employs a metric-enhanced feature transformation that maps an input vector $\mathbf{x} \in \mathbb{R}^d$ into a discriminative latent space $\mathbf{z} \in \mathbb{R}^r$ through a linear mapping

$$\mathbf{z} = \phi(\mathbf{x}) = \mathbf{L}\mathbf{x}, \qquad \mathbf{L} \in \mathbb{R}^{r \times d}. \tag{4}$$

The transformation matrix $\mathbf{L}$ encodes feature relevance and correlation structure, thereby reshaping the geometry of the input space into one that is more suitable for similarity computation and probabilistic density estimation.

This transformation induces a Mahalanobis-type distance that measures similarity after re-scaling and correlating the original features. Specifically, the distance between two samples $\mathbf{x}$ and $\mathbf{x}'$ is computed in the transformed space as

$$d_L(\mathbf{x}, \mathbf{x}') = \|\mathbf{L}\mathbf{x} - \mathbf{L}\mathbf{x}'\|_2, \tag{5}$$

where the linear mapping $\mathbf{L}$ projects the data into a metric-enhanced space in which feature relevance and correlations are explicitly encoded. Equivalently, this distance can be expressed in the original feature space as a Mahalanobis distance

$$d_M(\mathbf{x}, \mathbf{x}')^2 = (\mathbf{x} - \mathbf{x}')^\top \mathbf{M} (\mathbf{x} - \mathbf{x}'), \qquad \mathbf{M} = \mathbf{L}^\top \mathbf{L} \succeq 0, \tag{6}$$

where $\mathbf{M}$ is a positive semidefinite metric matrix. The diagonal elements of $\mathbf{M}$ control feature-wise scaling, while the off-diagonal elements capture inter-feature dependencies, enabling a more flexible and discriminative similarity measure than the isotropic Euclidean distance.

By learning an appropriate metric, the proposed transformation suppresses irrelevant dimensions, captures inter-feature dependencies, and yields a more stable and discriminative similarity structure for kernel-based density estimation.

B. Hybrid Kernel Design

Classical PNNs rely on a single Gaussian kernel to estimate class-conditional densities, which provides smooth local modeling but offers limited flexibility in capturing complex global

structures and heterogeneous data geometries. To enhance similarity expressiveness, the proposed MEHK-PNN introduces a hybrid kernel strategy that combines complementary kernel functions within a metric-enhanced feature space.

Specifically, let $\mathbf{z} = \mathbf{L}\mathbf{x}$ and $\mathbf{z}_i = \mathbf{L}\mathbf{x}_i$ denote the transformed representations of a test sample and a training sample, respectively. In this transformed space, the hybrid kernel integrates two complementary notions of similarity: local proximity captured by the Gaussian kernel and global relational structure captured by an additional kernel. By combining these components, the proposed approach is able to more effectively model complex and heterogeneous data distributions.

$$K_{\text{hyb}}(\mathbf{z}, \mathbf{z}_i) = \alpha\, K_G(\mathbf{z}, \mathbf{z}_i) + (1-\alpha)\, K_2(\mathbf{z}, \mathbf{z}_i), \qquad 0 \le \alpha \le 1, \tag{7}$$

where $K_G(\cdot,\cdot)$ is a Gaussian kernel ensuring stable local density estimation, $K_2(\cdot,\cdot)$ denotes a complementary similarity function, and α controls their relative contributions.

In this work, two configurations of the proposed hybrid kernel are investigated. The first combines the Gaussian kernel with a cosine similarity kernel, emphasizing directional consistency and scale invariance; this variant is referred to as MEHK-PNN-I. The second integrates the Gaussian kernel with a sigmoid kernel to introduce additional nonlinear discrimination capability, and is denoted as MEHK-PNN-II. These complementary formulations enable MEHK-PNN to jointly capture local neighborhood smoothness and broader relational structures within a unified similarity framework.

C. Hybrid Kernel Parzen Density Estimation

In MEHK-PNN, Parzen window density estimation is performed in the metric-enhanced feature space using the proposed hybrid kernels. For a given class c, the class-conditional density is estimated as

$$f_c(\mathbf{x}) = \frac{1}{n_c} \sum_{\mathbf{x}_i \in C_c} K_{\text{hyb}}(\mathbf{z}, \mathbf{z}_i), \tag{8}$$

where n_c is the number of training samples in class c and C_c denotes the corresponding index set.

By combining metric-aware distance computation with hybrid similarity modeling, the proposed Parzen estimator benefits from both improved geometric representation and enhanced expressive power. The Gaussian component provides smooth local density estimation, while the complementary kernel captures additional nonlinear or directional relationships between samples. As a result, MEHK-PNN yields more robust and discriminative class-conditional density estimates in high-dimensional and heterogeneous data environments.

Algorithm 1 Algorithm for the proposed Metric-Enhanced Hybrid Kernel Probabilistic Neural Network

Require: Training data $\mathcal{D} = \{(\mathbf{x}_k, y_k)\}_{k=1}^{N}$, metric transformation $\mathbf{L}$, hybrid weight α, Gaussian bandwidth σ, optional parameters for K_2, query sample $\mathbf{x}^*$
Ensure: Predicted label $\hat{y}$
Metric-enhanced transformation: $\mathbf{z}^* \leftarrow \mathbf{L}\mathbf{x}^*$
for $k = 1$ to N **do**
  $\mathbf{z}_k \leftarrow \mathbf{L}\mathbf{x}_k$
end for
Hybrid kernel construction:
$K_G(\mathbf{z}, \mathbf{z}') \leftarrow \exp\left(-\frac{\|\mathbf{z}-\mathbf{z}'\|_2^2}{2\sigma^2}\right)$
Define $K_2(\cdot,\cdot)$ as cosine similarity (MEHK-PNN-I) or sigmoid kernel (MEHK-PNN-II)
$K_{\text{hyb}}(\mathbf{z}, \mathbf{z}') \leftarrow \alpha\, K_G(\mathbf{z}, \mathbf{z}') + (1-\alpha)\, K_2(\mathbf{z}, \mathbf{z}')$
Class-conditional density estimation:
for $c = 1$ to C **do**
  $f_c(\mathbf{x}^*) \leftarrow \frac{1}{n_c} \sum_{k:y_k=c} K_{\text{hyb}}(\mathbf{z}^*, \mathbf{z}_k)$
end for
Decision rule:
$\hat{y} \leftarrow \arg\max_c f_c(\mathbf{x}^*)$

IV. Experiments & Discussion

This section evaluates the effectiveness of the proposed MEHK-PNN framework through comprehensive experiments on 35 benchmark KEEL datasets [21]. The proposed method is compared against several probabilistic and discriminative classifiers. These include the classical PNN [15] with a Gaussian kernel, its enhanced variants such as Improved PNN (I-PNN) [19] and Weighted PNN (W-PNN) [18], as well as Skew-N-PNN [20], which employs asymmetric kernel modeling. In addition, strong discriminative baselines are considered, including the SVM with an RBF kernel [9], Random Forest (RF) [11], and XGBoost (XGB) [12], to provide a comprehensive performance comparison on tabular data. The detailed experimental setup, along with the hyperparameter configurations, is provided in Section S.I of the supplementary material.

A. Performance Analysis

Table I summarizes the average F1-score, standard deviation, and average rank of the proposed MEHK-PNN variants and competing classifiers across 35 KEEL benchmark datasets. The detailed per-dataset results are reported in Table S.I of the supplementary material. Overall, the proposed methods consistently outperform both probabilistic and discriminative baselines in terms of classification effectiveness and stability. Among all compared models, the proposed MEHK-PNN-II achieves the highest average F1-score of 88.3558, followed by the proposed MEHK-PNN-I with an average F1-score of 87.2743. In contrast, the classical PNN and its variants yield noticeably lower average F1-scores, including 85.0586 for PNN, 83.6221 for I-PNN, and 85.0026 for W-PNN. This performance gap highlights the limitations of single-kernel Gaussian density estimation when similarity is computed directly in the original feature space. The skewness-based Skew-N-PNN achieves an average F1-score of 85.0072, which is comparable to Gaussian-based PNN variants but does not provide a consistent improvement across datasets. When compared with strong discriminative classifiers, such

TABLE I: Average F1-score, standard deviation, and ranking of the proposed MEHK-PNN and baseline classifiers on 35 KEEL benchmark datasets.

Model →	**PNN [15]**	**I-PNN [19]**	**W-PNN [18]**	**SVM [9]**	**RF [11]**	**XGB [12]**	**Skew-N-PNN [20]**	**MEHK-PNN-I†**	**MEHK-PNN-II†**
Average F1-Score	85.0586	83.6221	85.0026	85.5497	85.4562	85.3819	85.0072	87.2743	**88.3558**
Average Std. Dev.	7.69	7.69	7.64	7.47	7.07	7.88	6.46	6.26	**6.12**
Average Rank	5.5714	7.1286	5.6857	5.2286	5.6857	5.7857	4.6286	3.0857	**2.2**

† indicates the proposed models.

as SVM, RF, and XGBoost, which achieve average F1-scores of 85.5497, 85.4562, and 85.3819 respectively, both MEHK-PNN variants demonstrate clear performance advantages. In addition to higher mean performance, the proposed methods also exhibit improved stability across datasets. As reported in Table I, MEHK-PNN-II attains the lowest average standard deviation of 6.12, followed closely by MEHK-PNN-I with 6.26. These values are consistently lower than those of all baseline methods, suggesting that the proposed framework delivers more reliable and less variable performance across heterogeneous datasets. A ranking-based evaluation further reinforces these observations. The proposed MEHK-PNN-II achieves the best (lowest) average rank of 2.2, while the proposed MEHK-PNN-I ranks second with an average rank of 3.0857. All baseline methods obtain substantially higher average ranks, reflecting less consistent performance across datasets. This ranking analysis confirms that the proposed methods not only achieve superior average F1-scores but also maintain robust performance across diverse classification scenarios.

Overall, the results demonstrate that the integration of metric-enhanced feature transformation and hybrid kernel similarity modeling significantly improves both the accuracy and stability of probabilistic classification. These findings validate the effectiveness of the proposed MEHK-PNN framework for complex, high-dimensional, and heterogeneous datasets.

Further, to assess the statistical significance of the observed performance differences, a nonparametric rank-based significance analysis is conducted using the Friedman test, followed by the Nemenyi post-hoc procedure. The Friedman test [22] results are reported in Table II. In this study, $m = 9$ classification models are evaluated over $D = 35$ KEEL benchmark datasets. The computed Friedman statistic is $\chi_F^2 = 84.5127$, with the corresponding Iman-Davenport statistic $F_F = 14.6988$. At the 5% significance level, the critical value of the F-distribution with degrees of freedom $(m-1, (m-1)(D-1)) = (8, 272)$ is 1.9725. Since the obtained F_F value substantially exceeds this critical threshold, the null hypothesis that all classifiers achieve equivalent performance is rejected. This confirms that the differences observed in average F1-score among the compared methods are statistically significant. Following this, the Nemenyi post-hoc analysis is employed to identify pairwise differences between individual models. The results of this analysis are summarized in Table III. Based on the average ranks, MEHK-PNN-II achieves the best overall rank of 2.20 and is therefore used as the reference method. The critical difference (C.D.) at the 5% significance level is 2.03. As shown in Table III, the rank differences between MEHK-PNN-II and all baseline classifiers, including classical PNN variants, Skew-N-PNN, SVM, RF, and XGBoost, exceed the critical difference threshold. This indicates that the performance improvements achieved by MEHK-PNN-II over these methods are statistically significant. In contrast, the rank difference between MEHK-PNN-II and MEHK-PNN-I is 0.8857, which is below the critical difference. This suggests that although MEHK-PNN-II attains the best average rank, its advantage over MEHK-PNN-I is not statistically significant. Overall, the Friedman and Nemenyi analyses jointly confirm that the proposed MEHK-PNN framework delivers statistically significant performance gains over existing probabilistic and discriminative classifiers, while also exhibiting consistent behavior across diverse datasets. These results provide strong statistical evidence supporting the effectiveness and robustness of the proposed metric-enhanced hybrid kernel design.

We also investigate the influence of hyperparameter variations on classification performance of the proposed MEHK-PNN. The discussion is provided in Section S.II of the supplementary file.

TABLE II: Statistical significance analysis using Friedman test.

m	D	χ_F^2	F_F	$F((m-1),(m-1)(D-1))$	Significant difference (As per Friedman test)
9	35	84.5127	14.6988	1.9725	Yes

TABLE III: Statistical significance analysis using Nemenyi post-hoc test.

Model	Average rank	Rank difference	Significant difference (As per Nemenyi post-hoc test)
PNN [15]	5.5714	3.3714	Yes
I-PNN [19]	7.1286	4.9286	Yes
W-PNN [18]	5.6857	3.4857	Yes
SVM [9]	5.2286	3.0286	Yes
RF [11]	5.6857	3.4857	Yes
XGB [12]	5.7857	3.5857	Yes
Skew-N-PNN [20]	4.6286	2.4286	Yes
MEHK-PNN-I†	3.0857	0.8857	No
MEHK-PNN-II†	2.2	-	N/A

At the 5% significance level, the critical difference (C.D.) is 2.03.

V. Conclusions

This work introduced the Metric-Enhanced Hybrid Kernel Probabilistic Neural Network (MEHK-PNN), a robust probabilistic classification framework that enhances classical PNNs through metric-aware similarity modeling and hybrid kernel density estimation. The proposed approach preserves the analytical transparency and non-iterative learning paradigm of PNNs while substantially improving their representational and discriminative capabilities. Comprehensive experiments

on 35 KEEL benchmark datasets demonstrate that both variants of MEHK-PNN consistently outperform classical PNNs and competitive baselines in terms of F1-score and average rank. In particular, the MEHK-PNN-II variant, which incorporates a Gaussian-sigmoid hybrid kernel, achieves the best overall performance, indicating that introducing nonlinear global similarity alongside local density modeling provides a more expressive and discriminative Parzen estimator than the Gaussian-cosine formulation used in MEHK-PNN-I. Beyond higher mean performance, MEHK-PNN-II also exhibits lower standard deviation across cross-validation folds, reflecting improved stability and reduced sensitivity to data partitioning. Statistical significance analysis using the Friedman test followed by the Nemenyi post-hoc procedure confirms that the observed improvements are not incidental but statistically reliable across datasets. Furthermore, sensitivity analysis with respect to the Gaussian bandwidth and hybrid kernel weight reveals broad regions of stable high F1-score, demonstrating that MEHK-PNN maintains strong performance without requiring fine-grained hyperparameter tuning. This robustness is particularly desirable in practical scenarios where optimal parameter selection is challenging. In future work, adaptive hybrid kernel weighting and learnable metric transformations can be explored to further improve flexibility and reduce manual tuning. In addition, extending the proposed framework to large-scale problems through prototype selection or sparsification strategies represents a promising direction.

REFERENCES

[1] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[2] M. Akhtar, M. Tanveer, and M. Arshad, "RoBoTS: A robust bounded twin svm based on roboss loss function," *Pattern Recognition*, vol. 179, p. 113653, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320326006187

[3] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[4] M. Akhtar, M. Tanveer, M. Arshad, and for the Alzheimer's Disease Neuroimaging Initiative, "Advancing supervised learning with the wave loss function: A robust and smooth approach," *Pattern Recognition*, p. 110637, 2024. [Online]. Available: 10.1016/j.patcog.2024.110637

[5] M. Tanveer *et al.*, "Ensemble deep learning for alzheimer's disease characterization and estimation," *Nature Mental Health*, vol. 2, no. 6, pp. 655–667, 2024. [Online]. Available: https://doi.org/10.1038/s44220-024-00237-x

[6] A. Kumari, M. Akhtar, M. Tanveer, and M. Arshad, "Diagnosis of breast cancer using flexible pinball loss support vector machine," *Applied Soft Computing*, p. 111454, 2024. [Online]. Available: 10.1016/j.asoc.2024.111454

[7] M. U. Friedrich *et al.*, "Computer vision in clinical neurology: a review," *JAMA Neurology*, 2025. [Online]. Available: 10.1001/jamaneurol.2024.5326

[8] M. Tanveer, M. Sajid, M. Akhtar, A. Quadir, T. Goel, A. Aimen, S. Mitra, Y. D. Zhang, C. T. Lin, and J. D. Ser, "Fuzzy deep learning for the diagnosis of Alzheimer's disease: Approaches and challenges," *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 10, pp. 5477–5492, 2024.

[9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[10] M. Akhtar, M. Tanveer, and M. Arshad, "RoBoSS: A robust, bounded, sparse, and smooth loss function for supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–13, 2024. [Online]. Available: 10.1109/TPAMI.2024.3465535

[11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[12] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[13] I. Goodfellow *et al.*, *Deep learning*. MIT Press Cambridge, 2016, vol. 1, no. 2.

[14] V. Hassija *et al.*, "Interpreting black-box models: a review on explainable artificial intelligence," *Cognitive Computation*, vol. 16, no. 1, pp. 45–74, 2024.

[15] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.

[16] B. Mohebali *et al.*, "Probabilistic neural networks: a brief overview of theory, implementation, and application," *Handbook of Probabilistic Models*, pp. 347–367, 2020.

[17] Y. Zeinali and B. A. Story, "Competitive probabilistic neural network," *Integrated Computer-Aided Engineering*, vol. 24, no. 2, pp. 105–118, 2017.

[18] M. Kusy and P. A. Kowalski, "Weighted probabilistic neural network," *Information Sciences*, vol. 430, pp. 65–76, 2018.

[19] I. Izonin, R. Tkachenko, and M. Greguš, "I-PNN: An improved probabilistic neural network for binary classification of imbalanced medical data," in *Database and Expert Systems Applications*. Cham: Springer International Publishing, 2022, pp. 147–157.

[20] S. M. Naik *et al.*, "Skew-probabilistic neural networks for learning from imbalanced data," *Pattern Recognition*, vol. 165, p. 111677, 2025.

[21] J. Derrac *et al.*, "KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *J. Mult. Valued Log. Soft Comput*, vol. 17, pp. 255–287, 2015.

[22] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, no. Jan, pp. 1–30, 2006.