Available online at www.eccomasproceedia.org
Eccomas Proceedia UNCECOMP (2025) 43-55

ECCOMAS
Proceedia

UNCECOMP 2025
6th ECCOMAS Thematic Conference on
Uncertainty Quantification in Computational Science and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Rhodes Island, Greece, 15-18 June 2025

POSSIBILISTIC NEURAL NETWORKS: RELIABLE CONFIDENCE PREDICTIONS UNDER LIMITED DATA

Jan Schneider[1], **Tom Könecke**[1], **and Michael Hanss**[1]

[1]Institute of Engineering and Computational Mechanics
Pfaffenwaldring 9, 70569 Stuttgart
e-mail: {jan.schneider, tom.koenecke, michael.hanss}@itm.uni-stuttgart.de

Abstract. *Neural networks excel at capturing complex data relationships across diverse fields, yet they often lack the ability to quantify prediction uncertainties—a critical requirement for applications like medical diagnosis and autonomous driving. While existing solutions like Bayesian neural networks and interval-based methods address this issue, they typically demand substantial computational resources, large data sets, and often rely on stochastic assumptions that may not suit scenarios with both aleatoric and epistemic uncertainties. This paper introduces possibilistic neural networks (PNNs), a novel approach for providing reliable confidence predictions with limited data. Based on possibility theory, PNNs offer more flexible uncertainty modeling than traditional probabilistic methods while maintaining stochastic guarantees. The method employs a simple architecture of fully connected layers and introduces a novel three-part loss function to ensure nested confidence intervals of decreasing width. The network is trained to predict a plausibility function, which is then transformed into a possibility function using either approximate or reliable transformations. The effectiveness of the proposed approach is demonstrated through experiments with both Gaussian and non-Gaussian noise scenarios, using only* 80 *training samples. Results show that PNNs can successfully generate valid nested confidence intervals without requiring prior knowledge of the underlying distributions. The reliable transformation extends the method to provide guaranteed confidence predictions even with limited data sets, making it particularly suitable for real-world applications where data scarcity and uncertainty quantification are crucial concerns.*

Keywords: Possibility Theory, Neural Networks, Limited Data, Confidenc Prediction, Uncertainty Quantification

ISSN:2623-3339
Peer-review under responsibility of the organizing committee of UNCECOMP 2025.
doi: 10.7712/120225.12339.21200

1 INTRODUCTION

Machine learning, and particularly neural networks, have revolutionized numerous field by providing powerful tools for modeling complex relationships in data. From medical diagnosis to autonomous systems, these methods have demonstrated remarkable capabilities in pattern recognition and prediction tasks. However, as these applications increasingly impact critical decision-making processes, the need for reliable uncertainty quantificatio has become an area of major interest. Traditional neural network approaches focus primarily on minimizing prediction errors, resulting in point estimates that may appear deceptively precise. This limitation becomes particularly problematic in safety-critical applications, where understandingthe confidenc level of a prediction is as crucial as the prediction itself. Furthermore, real-world applications often face the dual challenge of limited training data and the presence of both aleatoric and epistemic uncertainties.

There are various approaches to capture uncertainty in machine learning tasks and neural networks using interval arithmetic, Bayesian neural networks, or robust training based on Lipschitz bounds [10, 8, 9, 3]. These neural networks can be trained on noisy data and capture underlying relations to outputs which enables quantifie statements in the prediction step. However, these methods often require prior assumptions or information on the underlying noise distribution, prior guesses on appropriate basis functions or they suffer from a complex network structure.

In contrast, we propose a novel approach for neural network training with uncertain data based on possibility theory. Possibility theory, as a theory of imprecise probabilities, grants a fl xible tool to describe uncertainty without the need of prior assumptions. This allows to model aleatoric and epistemic uncertainty sufficientl while still being able to provide stochastical guarantees. The proposed method, called Possibilistic Neural Networks (PNNs), is designed to provide reliable confidenc predictions for regression tasks with limited data. The network is trained to predict a plausibility function which is then transformed into a possibility function by the use of either either approximate or reliable transformations. The main contribution of this paper is the derivation of a suitable cost function for the neural network training. The effectiveness of the proposed approach is demonstrated in scenarios with both Gaussian and non-Gaussian noise, using only a limited number of training samples. The experiments demonstrate the method's ability to construct valid nested confidenc intervals, even without knowing the distributions of the input data beforehand. By incorporating a reliable transformation technique, this approach remains effective even when working with small data sets. This makes PNNs especially valuable for practical applications where both limited data availability and the need to accurately measure uncertainty pose significan challenges.

The remainder of this paper is organized as follows: Section 2 provides the theoretical background of possibility theory and its application as a tool for uncertainty quantifi cation. Section 3 introduces the architecture and training methodology of PNNs and states the loss function as well as the analytical gradients for the network training. Section 4 presents experimental results on both Gaussian and non-Gaussian noise scenarios, demonstrating the method's effectiveness. Finally, Section 5 discusses the implications and potential applications of this approach, along with directions for future research.

2 POSSIBILITY THEORY

In the following paragraph, a brief introduction to possibility theory is given that only provides the essential background for understanding the proposed neural-network learning scheme and the idea of confidenc predictions. This introduction does not aim to be exhaustive and for a more detailed discussion the reader is referred to [5, 7].

Consider a classical random experiment with the sample space Ω and its associated σ-algebra Σ. For an event $U \in \Sigma$ with counter event $\neg U$, the so-called elementary possibility function is any measurable function $\pi : \Sigma \to [0, 1]$ satisfying

$$\sup_{\omega \in \Omega} \pi(\omega) = 1. \tag{1}$$

Subsequently, this allows to introduce the possibility measure

$$\Pi(U) = \sup_{\omega \in U} \pi(\omega) \; \forall \; U \in \Sigma, \tag{2}$$

which is induced by the elementary possibility function and satisfie the following properties:

$$\Pi(\emptyset) = 0, \quad \Pi(\Omega) = 1, \quad \Pi(U \cup V) = \max\{\Pi(U), \Pi(V)\}. \tag{3}$$

There exist different characteristics of an elementary possibility function to quantify the information with respect to a probabilistic statement. One key concept is consistency [2, 1], which states that a probability measure P is consistent with an elementary possibility function if the probability $\mathrm{P}(U)$ is upper-bounded by the induced possibility measure $\Pi(U)$ for all events $U \in \Sigma$. This concept forms the foundation for understanding the later proposed statement relating possibility functions to probabilities and, thus, to confidenc predictions. An essential concluding property arises in the context of strict superlevel sets of the elementary possibility function, given by

$$C_\pi^\alpha = \{\omega \in \Omega \mid \pi(\omega) > \alpha\}. \tag{4}$$

These sets are also called confidenc sets and specify regions in the sample space where the possibility measure exceeds a given threshold $\alpha \in [0, 1]$.

Subsequently, when an elementary possibility function is consistent with a probability measure, the confidenc sets can be used to derive the probabilistic guarantee from the possibility function via

$$P(C_\pi^\alpha) \geq 1 - \alpha. \tag{5}$$

In the context of inferential models [4], i.e., when the distribution provides predictive information about an unknown state, this property has to be assessed and the corresponding model is called valid if it holds for the whole sample space. This formulation is the basis for the later proposed neural-network learning scheme, which aims to predict confidenc sets for input instances. The assessment of validity provides a more intuitive understanding of possibility functions and facilitates comparisons between them. In fact, there are infinit possibility functions consistent with a probability measure. This motivates the introduction of specificit , which states that a possibility function π_1 is more specifi than another possibility function π_2 if $\pi_1(\omega) \leq \pi_2(\omega)$ for all $\omega \in \Omega$. Using the interpretation of Eq. (5), it is trivial to conclude that a more specifi possibility function yields a more informative result since the confidenc sets are tighter, i.e. $C_{\pi_1}^\alpha \subseteq C_{\pi_2}^\alpha$ for all $\alpha \in [0, 1]$.

These concepts and characteristics extend to quantitative statements about imprecise probabilities through the introduction of a subjective plausibility function $\rho : \Sigma \to [0, 1]$, which quantifie event plausibility based on available information. While ρ is not subjected to the constraints of a possibility function and, technically, does not have to map on the same metric, it must establish a relative ordering of event plausibility. Therefore, given a family $\mathcal{P}$ of probability measures, an appropriate plausibility function can be used to defin the plausibility-to-possibility- or IP-Π-Transform [5]

$$\pi(\omega) = \sup_{P \in \mathcal{P}} P\{\xi \in \Omega : \rho(\xi) \leq \rho(\omega)\} \quad \forall\, \omega \in \Omega. \tag{6}$$

To this end, one of the crucial challenges is to identify an appropriate plausibility function. While existing approaches for data-driven applications rely on prior knowledge of the data-generating process, we propose a neural-network-based approach to learn plausibility functions directly from data for predicting instance-specifi confidenc sets. Furthermore, the vaildity of the neural-network-based predictions is of utmost importance to guarantee reliable predictions on unseen data.

3 NEURAL NETWORKS AS PLAUSIBILITY MEASURES

Several approaches have been proposed to capture uncertainty in training data for reliable predictions of neural networks [10, 8]. However, these methods often suffer from complex structures or require prior assumptions. The approach proposed in this paper is a quantitative learning scheme using a classical neural network that predicts plausibility

functions, which are then used to derive confiden e sets and, subsequently, possibility functions.

The general neural network architecture is intentionally kept simple. It consists of an input layer with one neuron per input dimension, followed by a series of l hidden layers, each containing k_l neurons. Each hidden neuron $h^j_{k_j}$ is fully connected to either the successor layer or the output layer. Finally, the output layer contains $2n$ neurons, with corresponding outputs $\boldsymbol{y} = y_1, \ldots, y_{2n}$, where $n \in \mathbb{N}$ represents the desired number of confidenc intervals. Each neuron is structured classically as a linear combination of weighted inputs and bias terms, which are fed to a specifie activation function. The network is trained on N pairs of data points, consisting of input instances $x \in \mathbb{R}^d$ and noisy output values $\hat{y} \in \mathbb{R}$, which can be interpreted as realizations of a random process. The general structure is visualized in Figure 1. The key idea lies in the interpretation of

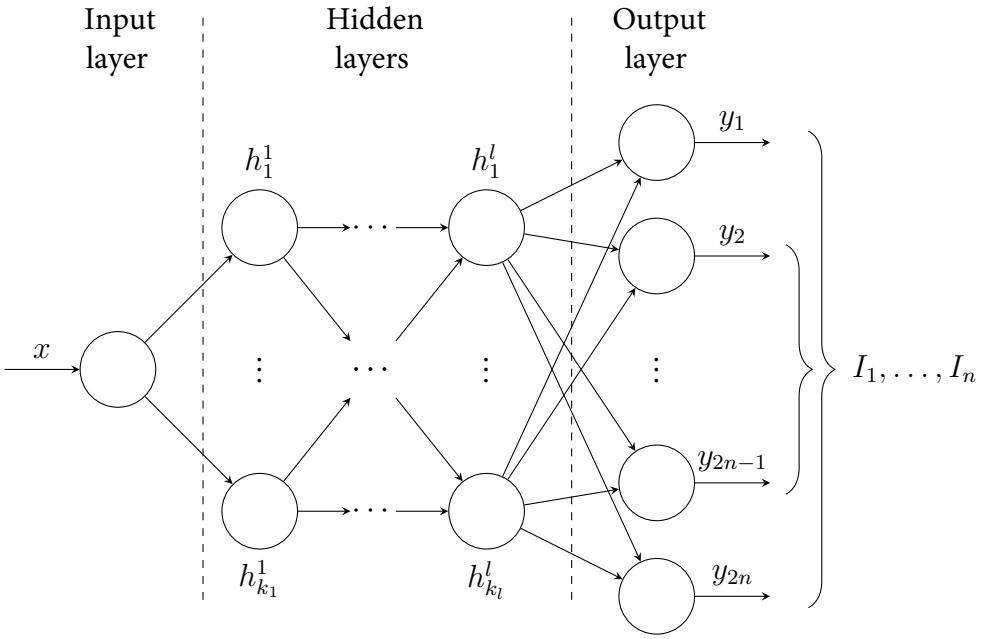


Figure 1: General structure of the neural network with interval outputs.

the outputs: every i th and $(2n - i + 1)$ th neuron in the output layer defin one prediction interval I_i. The challenge consists in developing an appropriate loss function that ensures these intervals to provide a nested plausibility ordering while enabling maximal specificit and validity of the transformed possibility functions.

For the proposed method, consider a three-part loss

$$L(\boldsymbol{y}, \hat{y}) = \boldsymbol{\lambda} \begin{bmatrix} L_{\mathrm{w}}(\boldsymbol{y}) & L_{\mathrm{o}}(\boldsymbol{y}) & L_{\mathrm{n}}(\boldsymbol{y}, \hat{y}) \end{bmatrix}^\top , \tag{7}$$

where $L_\mathrm{w}, L_\mathrm{o}, L_\mathrm{n}$ are individual loss terms for different optimization goals weighted by respective hyperparameters $\boldsymbol{\lambda} = [\lambda_\mathrm{w}, \lambda_\mathrm{o}, \lambda_\mathrm{n}]$.

The firs term L_w penalizes the width of the prediction interval. As stated earlier, the goal is to be as specifi as possible since boundaries over the whole state space would still be valid but contain no information. Since confidenc intervals with high α-values should be smaller than lower ones, the partial loss of the corresponding interval $I_i = [y_i, y_{2n-i+1}]$ is scaled with an interval dependent factor. The complete loss term is given by

$$L_\mathrm{w}(\boldsymbol{y}) = \sum_{i=1}^{n} \left(\frac{i}{2n} + 1 \right) |y_{2n-i+1} - y_i|. \tag{8}$$

For backpropagation, the gradient of this partial loss function has to be calculated with respect to each lower interval bound y_i and upper interval bound y_{2n-i+1} independently. The gradients for the successor layers are then calculated by the chain rule. Analytically, this can be derived as

$$\frac{\delta L_\mathrm{w}}{\delta y_i} = \left(\frac{i}{2n} + 1 \right) \mathrm{sgn}(y_{2n-i+1} - y_i), \quad \frac{\delta L_\mathrm{w}}{\delta y_{2n-i+1}} = -\frac{\delta L_\mathrm{w}}{\delta y_i}. \tag{9}$$

Furthermore, the second term L_o is responsible for ordering the intervals. It is trivial to see that the inequality $y_i \leq y_{2n-i+1}$ has to hold for all I_i with $i \in \{1, \ldots, n\}$. Besides, it is important to ensure a hierarchical ordering between the intervals, i.e., the lower bound of the $(i+1)$ th interval has to be greater than (or equal to) the upper bound of the i th interval. These relations can be enforced by definin the second loss term as

$$L_\mathrm{o}(\boldsymbol{y}) = \frac{1}{2n} \sum_{j=1}^{2n-1} \max(0, y_{j+1} - y_j). \tag{10}$$

Note that it is sufficien to enforce $y_{j+1} \leq y_j$ for all j since this implies $y_{j+2} \leq y_j$. Analogously to the width loss, the gradient of this partial loss function with respect to the outputs has to be calculated independently for each node with

$$\frac{\delta L_\mathrm{o}}{\delta y_1} = -\frac{\mathbb{1}_{\mathbb{R}^+}(y_2 - y_1)}{2n}, \quad \frac{\delta L_\mathrm{o}}{\delta y_{2n}} = \frac{\mathbb{1}_{\mathbb{R}^+}(y_{2n} - y_{2n-1})}{2n}, \tag{11}$$

$$\frac{\delta L_\mathrm{o}}{\delta y_j} = \frac{-\mathbb{1}_{\mathbb{R}^+}(y_{j+1} - y_j) + \mathbb{1}_{\mathbb{R}^+}(y_j - y_{j-1})}{2n}, \quad j \in \{2, \ldots, 2n-1\}, \tag{12}$$

where $\mathbb{1}_{\mathbb{R}^+}$ is the indicator function for positive real numbers.

Lastly, the third term L_n handles prediction errors by penalizing intervals that do not contain the true value $\hat{y}$. When the true value falls outside an interval, this loss term encourages two corrective actions: widening the interval and shifting its center towards $\hat{y}$.

However, not all intervals need to provide the same level of confidence For instance, a 99% confidenc interval should almost always contain the true value, while a 50% confidenc interval can reasonably exclude it about half the time.

This motivates the introductionof the loss term

$$L_{\mathrm{n}}(\boldsymbol{y}, \hat{y}) = \frac{1}{2N} \sum_{i=1}^{n} \left(\frac{2n}{i} \right)^2 \left([\max(0, \hat{y} - y_i)]^2 + [\max(0, y_{2n-i+1} - \hat{y})]^2 \right), \quad (13)$$

where the hierarchical confidenc structure is implemented through the quadratic scaling factor $((2n)/i)^2$. For intervals meant to provide high confidenc (small i relative to n), this factor creates a large penalty when the true value falls outside the interval. Conversely, for intervals meant to provide lower confidenc (large i relative to n), the penalty is reduced, allowing these intervals to remain narrower even if they occasionally exclude the true value.

The gradient for the currently considered interval can again be derived independently as

$$\frac{\delta L_{\mathrm{n}}}{\delta y_i} = -\frac{1}{N} \left(\frac{2n}{i} \right)^2 \max(0, \hat{y} - y_i), \quad (14)$$

$$\frac{\delta L_{\mathrm{n}}}{\delta y_{2n-i+1}} = \frac{1}{N} \left(\frac{2n}{i} \right)^2 \max(0, y_{2n-i+1} - \hat{y}). \quad (15)$$

Finally, the total gradient

$$\frac{\delta L}{\delta y_i} = \lambda_{\mathrm{w}} \frac{\delta L_{\mathrm{w}}}{\delta y_i} + \lambda_{\mathrm{o}} \frac{\delta L_{\mathrm{o}}}{\delta y_i} + \lambda_{\mathrm{n}} \frac{\delta L_{\mathrm{n}}}{\delta y_i} \quad (16)$$

can be calculated for each output neuron allowing to determine the gradients of each hidden neuron accordingly. After training, the outputs of the network defin a plausibility function $\rho_Y(y)$. For n total confidenc intervals, the plausibility of each interval is define by $\rho_{Y,I_i} = i/n$ with $i \in \{1, \ldots, n\}$.

To obtain a possibility function consistent with the plausibility function, the plausibility-to-possibility transform must be applied. However, since only limited data are available for training and transformation, different approaches can be considered. The firs approach is the approximate transformation [5], which define the unbiased estimator

$$\bar{\pi}_Y(y) = \frac{K(y)}{N}, \quad K(y) = |\{Y_k : \rho_Y(Y_k) \leq \rho_Y(y); k = 1, \ldots, N\}|, \quad \forall y \in Y, \quad (17)$$

where $K(y)$ counts the number of samples with plausibility less than or equal to $\rho_Y(y)$. As shown in [6], this counting mechanism provides a sufficien statistic for the transformation. However, while this approach maintains validity on the training data set,

it cannot provide guaranteed coverage for unseen data. Hence, as stated in [6], a reliable transformation can be applied using the inverse of the beta cumulative distribution function beta^{-1}, resulting in

$$\hat{\pi}_Y(y) = \text{beta}^{-1}(\gamma, K(y)+1, N-K(y)), \quad \forall y \in Y, \tag{18}$$

where the reliability level $\gamma \in [0,1]$ states that $\hat{\pi}_Y(y)$ is point-wise less specifi than $\pi_Y(y)$ with a probability of

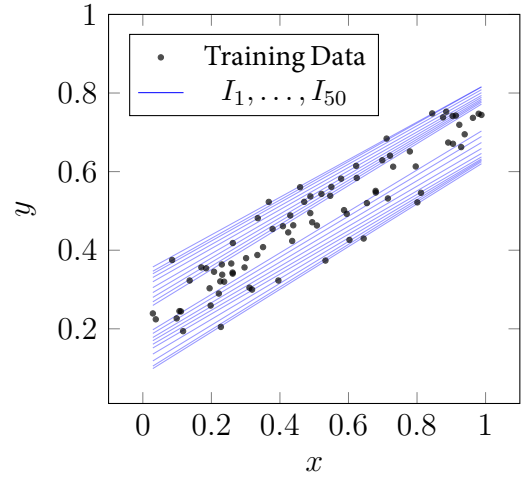
$$P\left(\hat{\pi}_Y(y) \geq \pi_Y(y)\right) \geq \gamma. \tag{19}$$

With the reliable extension, it is possible to loosen the specificit learned by the network on the training data set and generalize the prediction to unseen data while still being able to give guarantees on the predictions. The γ-level define a trade-off parameter between specificit and conservatism and can be tuned to the desired level of reliability.

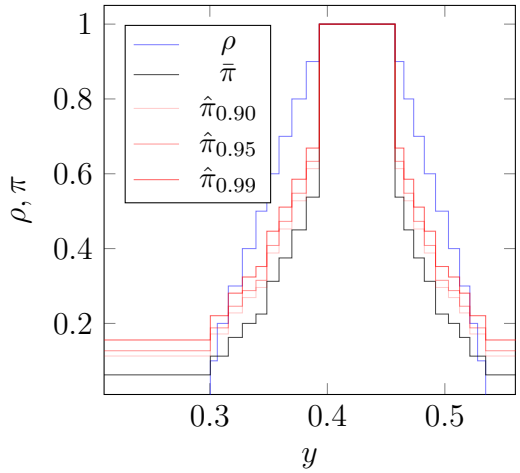
4 RESULTS

To verify the proposed method, two neural networks are trained for different test cases. As a firs example, consider a data set drawn from a linear function $\hat{y} = 0.2x+1+w$ with a Gaussian distributed noise term $w \sim \mathcal{N}(0, 0.1)$ to simulate aleatory measurement noise. The training set comprises $N = 80$ randomly sampled values of $x \in [0,5]$ and the corresponding noisy measurements $\hat{y}$ which are normalized to the interval $[0,1]$. Both networks share the same architecture of two hidden layers with $k_1 = 2$ and $k_2 = 6$ neurons using the ReLu activation function, but differ in their output layers: the firs network has $m = 20$ neurons to predict $n = 10$ confidenc intervals, while the second has $m = 100$ outputs to predict $n = 50$ intervals. In their output layers, both networks employ linear activation functions and undergo training for 25.000 epochs with a learning rate of 0.0008. Since the networks performance is highly sensitive to hyperparameter selection, a grid-based optimization procedure is employed, yielding optimal hyperparameters $\boldsymbol{\lambda} = [0.0005, 0.001, 1.5]$ for the firs network and $\boldsymbol{\lambda} = [0.001, 0.003, 1.5]$ for the second network. As shown in Figure 2(a) and Figure 3(a), the networks are able to produce confidenc intervals which are nested, have a decreasing width and bound the data set tightly. In order to achieve a possibilistically valid result, $K(y)$ is determined through a post-processing analysis that quantifie the number of training data points contained within each interval. This allows to derive the approximate and reliable possibility function, which use the same interval boundaries but with respective possibilistic values. The differences can be observed when considering the prediction on a specifi x-value. The results for the approximate and reliable transformation with different γ-levels at $x = 0.4$ are shown in Figure 2(b) and Figure 3(b).

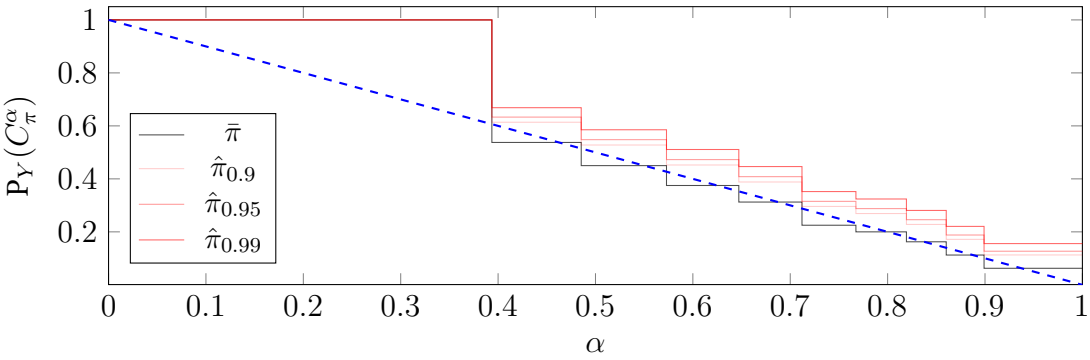
After verifying the nominal performance on the training data set, the next step is to analyze the network's predictive capabilities on unseen data and compare the reliable and approximate transformations. To do so, the consistency of confidenc predictions



(a) Network prediction ρ.



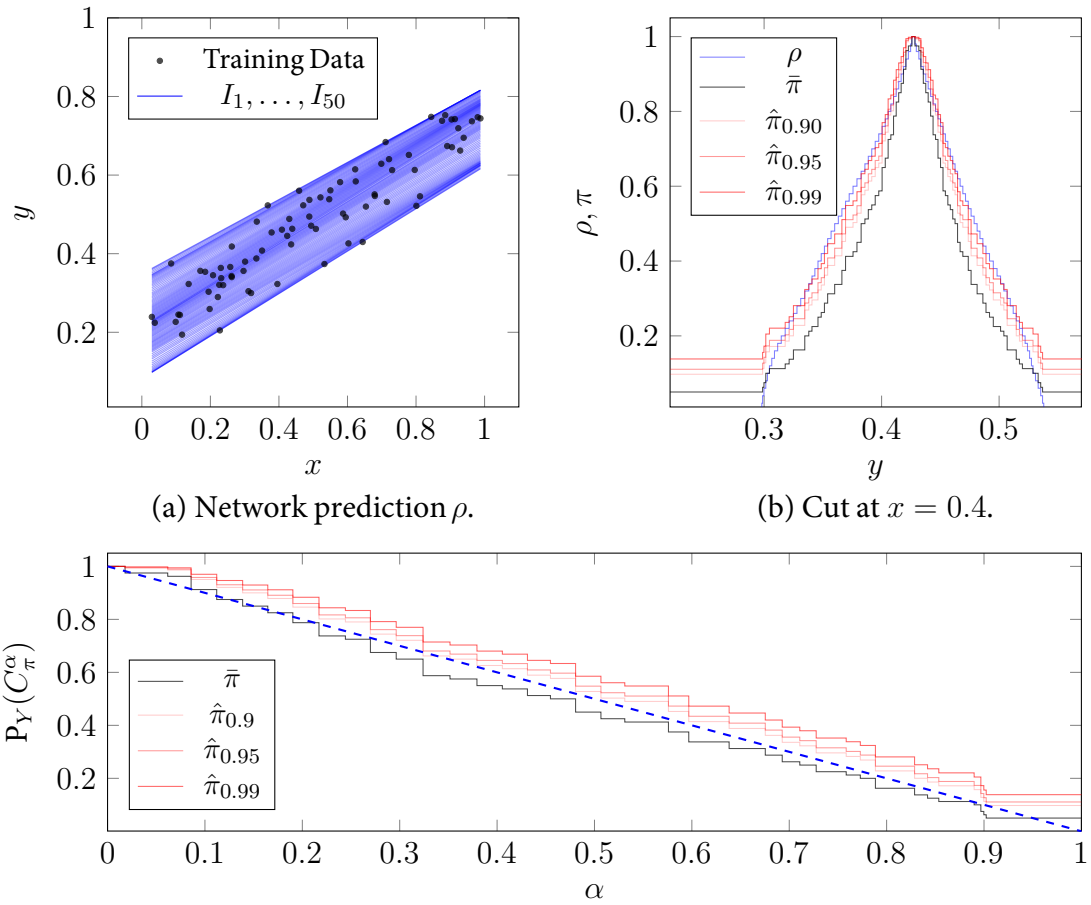
(b) Cut at $x = 0.4$.



(c) Evaluation of $P_Y(C_\pi^\alpha)$.

Figure 2: Prediction and validity results of network with ten confidenc intervals.

is evaluated on an unseen data set by checking Eq. (5) for different transformations. When assessing the number of correct predictions per interval, the corresponding graphs should remain above the $(1 - \alpha)$-line. This indicates that for any arbitrary α-level, the network is overly conservative but does not produce false predictions. The results for the same networks are shown in Figure 2(c) and Figure 3(c). As expected, the reliable transformation with increasing γ-levels is more conservative than the approximate transformation and allows to provide confiden predictions for unseen data. Note that with an increasing number of intervals, the possibility functions start to converge to the result of the Optimal Transform [4] of a normal distribution, which implies that the network is able to capture the underlying relationships between the probabilistic and possibilistic space.



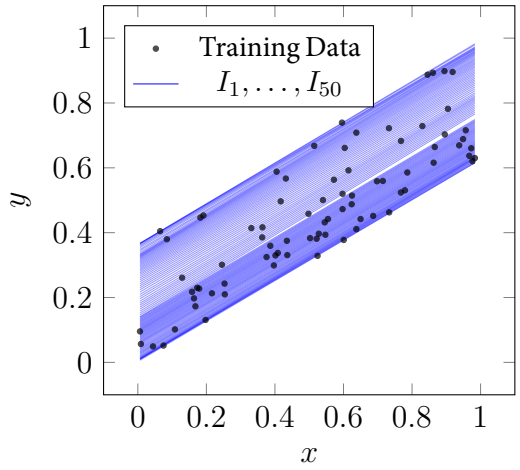
(a) Network prediction ρ.

(b) Cut at $x = 0.4$.

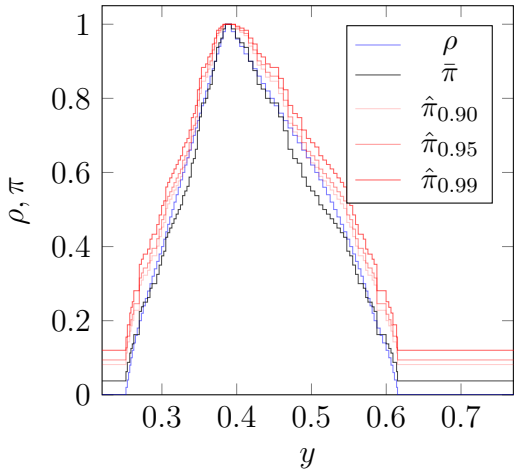
(c) Evaluation of $P_Y(C_\pi^\alpha)$.

Figure 3: Prediction and validity results of network with fift confidenc intervals.

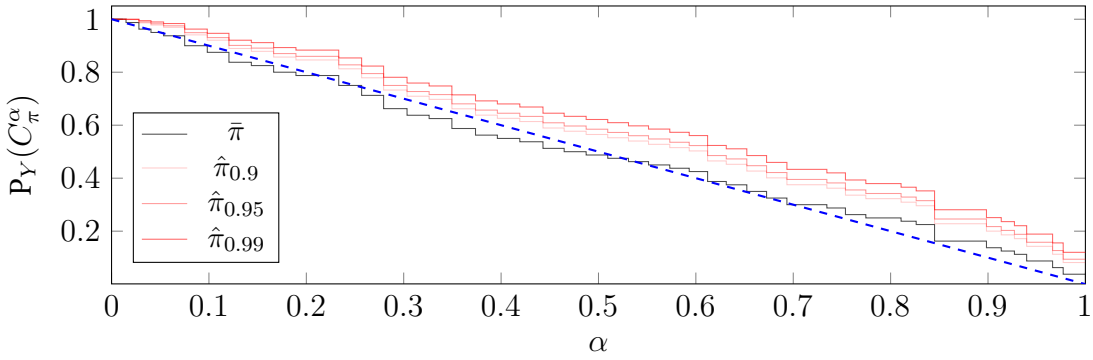
Furthermore, the great benefi of using this neural network structure under limited data is shown in a second example, where the network is trained on a data set which is influence by two combined uniform noise terms $W_1 \sim \mathcal{U}(-0.2, 0)$ and $W_2 \sim \mathcal{U}(0, 0.4)$. The random variable w is drawn from W_1 for one half of the total samples and from W_2 for the remaining half. As in the previous example, $N = 80$ randomly sampled input values x with their corresponding noisy measurements $\hat{y}$ are used. The same network architecture is maintained, but the hyperparameters are now optimized to $\boldsymbol{\lambda} = [0.0021, 0.003, 1.5]$. The results for the network prediction and the reliable and approximate transformationare shown in Figure 4. The network demonstrates that comparable performance to the Gaussian noise case can be achieved. This shows that the proposed method is able to provide reliable confidenc predictions under limited data and without any prior information about the underlying distribution.



(a) Network prediction ρ.



(b) Cut at $x = 0.4$.



(c) Evaluation of $P_Y(C_\pi^\alpha)$.

Figure 4: Prediction and validity results of network with fift confidenc intervals with uniform noise distribution.

5 CONCLUSION

In this work, a novel framework for uncertainty-aware neural-networks using possibility theory is introduced. The proposed methodology successfully quantifie uncertainty in training data by giving nested prediction intervals for possibilistic interpretations. While the main contribution is from the derivation of a suitable loss function for such specifi desires, the trained output can be extended by using the approximate and reliable transformation principles found in possibility theory. Especially the latter is meant to give reliable confidenc predictions on an unseen data set.

Empirical results demonstrate constant performance across diverse data sets, even for non-Gaussian noise distributions. It is especially remarkable that the network does not need any prior information or assumptions about the underlying distribution. While such characteristics show promise for applications in safety-critical environments like autonomous driving, challenges remain in capturing highly-nonlinear and/or multidimensional data distributions. Future research could be directed towards improving the convergence behavior and extending the proposed methodology, especially the sensitivity and optimization of the hyperparameters, which, so far, is the most challenging part for more complex data sets. One promising approach is to formalize different loss terms explicitly in the network architecture to reduce the sensitivity and number of hyperparameters while keeping the same guarantees. It is especially challenging to guarantee nested intervals over the whole domain rather then only on the training set, which could be resolved by this enhancement.

In general, possibilistic neural networks provide a promising alternative to traditional probabilistic methods, offering a principled approach to uncertainty representation that supports the development of more reliable AI systems.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support provided by the German Research Foundation(DFG) in the framework of the research project no. 514188140 (HA 2798/16-1). The authors would furthermorelike to express their sincere appreciation of the work contributed by Niklas Abraham in the framework of his student thesis under supervision of the authors.

References

[1] Miguel Delgado and Serafín Moral. "On the Concept of Possibility-Probability Consistency". In: *Fuzzy Sets and Systems* 21.3 (1987), pp. 311–318.

[2] Didier Dubois and Henri Prade. "When Upper Probabilities Are Possibility Measures". In: *Fuzzy Sets and Systems* 49.1 (1992), pp. 65–74.

[3] David Heckerman. *A Tutorial on Learning With Bayesian Networks*. 2022. arXiv: 2002.00269 [cs].

[4] Dominik Hose. "Possibilistic Reasoning with Imprecise Probabilities: Statistical Inference and Dynamic Filtering". In: *Schriften Aus Dem Institut Für Technische Und Numerische Mechanik Der Universität Stuttgart*. Dissertation, Band 74. Shaker Verlag, 2022.

[5] Dominik Hose and Michael Hanss. "A Universal Approach to Imprecise Probabilities in Possibility Theory". In: *International Journal of Approximate Reasoning* 133 (2021), pp. 133–158.

[6] Dominik Hose and Michael Hanss. "On Data-Based Estimation of Possibility Distributions". In: *Fuzzy Sets and Systems* 399 (2020), pp. 77–94.

[7] George Klir and Mark Wierman. *Uncertainty-Based Information: Elements of Generalized Information Theory*. Vol. 15. Springer Science & Business Media, 1999.

[8] Luis Oala et al. *Interval Neural Networks: Uncertainty Scores*. 2020. arXiv: 2003.11566.

[9] Patricia Pauli et al. "Training Robust Neural Networks Using Lipschitz Bounds". In: *IEEE Control Systems Letters* 6 (2022), pp. 121–126.

[10] Krasymyr Tretiak, Georg Schollmeyer, and Scott Ferson. "Neural Network Model for Imprecise Regression with Interval Dependent Variables". In: *Neural Networks* 161 (2023), pp. 550–564.