# Hierarchical Bayesian Neural Networks (HBNNs) for Large Data Classification

Mohsin Sami[1], Rabia Tehseen[1], Usman Aamir[1], Ramsha Saeed[1]
University of Central Punjab, Lahore, Pakistan

**Abstract**

Incorporating hierarchical structures into Bayesian frameworks gives rise to Hierarchical Bayesian Neural Networks (HBNNs), which enhance existing Bayesian neural networks by being able to model uncertainty in more complicated data. The present systematic literature review examines the evolution, use, and effectiveness of HBNNs pertaining to the training of large-scale classification tasks in different domains. The review aims to identify fully fledged modelling frameworks, discuss implementations involving real-world datasets, analyse the inference strategies and frameworks, and compare the relative performance to Bayesian and non-Bayesian models within the established frameworks. The review also highlights the gaps and issues of scaling HBNNs with the intent for big data classification. This study analyses 42 peer-reviewed publications dated between 2005 and 2025 and uncovers that HBNNs, in comparison to traditional models, consistently estimate uncertainty and reason probabilistically at better rates, but the lack of refinement in their parallel implementation due to high computational demand and convergence problems leads to a suboptimal solution for expansion bottlenecks. There is optimistic outlook with increasing focus on hybrid models, variational inference approaches, and spatially-aligned hierarchical priors for high-dimensional frameworks. This review provides the basic building blocks for researchers and practitioners seeking to harness the full potential of HBNNs in large-scale classification problems.

**Keywords:** Hierarchical Bayesian network, Deep Neural network, Bayesian Theory, convergence.

## 1. Introduction

The enormous amount of data being produced and stored by the healthcare and finance industry and numerous other allied autonomous systems have triggered a need for fast and precise machine learning algorithms. Though traditional deep neural networks (DNNs) have been documented to excel in large-scale classification tasks, their inability to measure uncertainty limits them in scenarios where high-stakes decisions need to be made, i.e., predictive uncertainty [1], [2]. This gap can be filled with Bayesian Neural Networks (BNNs) which provides a robust framework for modelling uncertainty such as the ones in network weights which enhances generalization and robustness [3].

Theoretically, BNNs provide a considerable advantage, but problems regarding their scalability, data storage inefficiencies and redundant data retrieval tend to arise. Methods like Bayes by Backprop [4] and Monte Carlo Dropout [5] which use shells and variational approximation to tackle these issues, but such methods cannot capture the hierarchical arrangement of the true data and tend to lose accuracy. To address the shortcomings of these methods, hierarchical structures of BNNs, or HBNNs, have been proposed. HBNNs facilitate further flexible and data driven regularization by placing hierarchical priors over parameters of the networks to model higher order uncertainty [6]. Currently available information shows that the ordered structure in question enhances posterior distributions in addition to improving large-scale transfer learning as well as domain adaptation [7].

The current research effort establishes the HBNN architecture that can be scaled and is configured to handle the classification of large datasets. We employ certain variational inference techniques in the form of amortized and structured variational family inference, which, while being economically efficient, do not compromise on the accuracy of the model. Through thorough benchmarking, we provide evidence that our method achieves a higher level of predictive performance with better uncertainty calibration and lower uncertainty as compared to the BNNs, and the deterministic DNNs.

The current research effort has been structured to include the following sections: In Section 1, we provide an overview of HBNN systems; in Section 2, we provide an overview of relevant literature; Section 3 deals with the description of the methods employed in this research; Section 4 is devoted to the presentation of the results, while Section 5 consolidates this report and proposes new research avenues.

## 2. Background of the study

The vast amounts of data and their increasing variety and velocity create some challenges for machine learning algorithms, classification algorithms in particular. Whereas many fields benefit from traditional deep learning algorithms, they still overfit the data, and lacking interpretability and poor quantification of uncertainty, they struggle with large-scale high-dimensional datasets. In recent times, to overcome these constraints, scholars have considered BNNs which allow for a probabilistic approach to both learning and inference. Of the many BNNs extensions, HBNNs are also a strong candidate, as they marry the representational capacity of deep neural models and the uncertainty of the inference mechanism of Bayesian models, but refined through the use of hierarchy [8]–[10].

Bayesian approaches leverage Bayes' theorem to quantify uncertainty in a model's parameters by considering them to be random variables and assigning them particular probability distributions. Forget the traditional point estimate learning mechanism; in BNNs, the weights of a neural network are represented by a distribution and predictions are made considering both aleatoric and epistemic uncertainties. Confidence in a model's predictions becomes vital especially in high-stake areas such as self-driving vehicles, finance, and medicine. These domains have the most profound impact on high-importance fields like self-driving cars, finance, and medicine, where understanding the confidence a model provides in its prediction is as significant as the prediction itself [11-12].

Nonetheless, standard BNNs tend to exhibit the most unappealing "feature" of inefficiency in terms of scalability, especially in the presence of large datasets. Performing Bayesian inference remains a very costly operation to carry out, especially with the parameters of deep learning models, which are inherently high-dimensional. Posterior distribution approximation through variational inference or Monte Carlo sampling techniques can be executed; however, these techniques often oversimplify uncertainty [13].

This is where HBNNs gain importance after HBNNs. By using hierarchical priors, an extra layer of stochasticity was introduced on top of the Bayesian framework. In a typical HBNN, instead of assigning fixed priors to the Model Parameters, a more flexible approach is used whereby hyperparameters are probabilistically governed by surrogates/hyperparameters. This specific modelling style enables the structure to be highly flexible and adaptable because it can learn prior distributions from the data using a data-driven approach [14], [15]. This didactic approach is especially effective for large-scale classification problems where prior information is sparse and/or noisy.

The idea of hierarchy in Bayesian modelling is not novel and has been successfully used in other domains such as natural languages, imaging as well as in time-series analysis. Unlike Bayesian models, HBNNs are able to capture heterogeneities and complicated dependencies far more effectively due to the use of neural networks. Hierarchical priors are able to differentiate between inter-class distinctions that overlap by supporting statistical sharing across associated features or classes. This can be useful in cases when subtler distinctions are vital, like a multi-class classification problem with overlapping categories [16].

The last couple of years have seen an increase in the availability of efficient computational resources which, coupled with scalable inference algorithms, has made the use of HBNNs more practical. Stochastic Variational Inference (SVI) and Hamiltonian Monte Carlo (HMC) or other sampling techniques have also been modified to be compatible with hierarchical models in deep learning. Furthermore, the emergence of probabilistic programming languages like Pyro, Edward2, and TensorFlow Probability has made it easier to build and train HBNNs on large datasets [17], [18]. The above mentioned resources have significantly lessened the barriers facing researchers and practitioners who wish to employ Bayesian approaches to deep learning and its real world applications.

Moreover, the combination of transfer learning, federated learning, and active learning expands the applicability of HBNNs to data abundant scenarios with special considerations to privacy. As [19] mentions, the Bayesian perspective enables local models to keep customized priors while sharing in a global posterior and helps federated settings slice metadata without compromising accuracy or privacy. The same active learning scenarios can benefit from HBNNs where the models can effectively quantify uncertainty to strategically sample from the data, thus significantly improving performance for a given budget of labeled data.

Given the advantages of HBNs, these approaches bring about some issues regarding the classification of large data volumes. Such issues encompass convergence, over-sensitivity to the specified prior, and the interpretability of the learned hierarchical structure being the most problematic and consequently the most overlooked. Additionally, a relative lack of work can be noted, particularly in contrast to contemporary HBNs using benchmarking datasets, further emphasizing the gap in comprehensive work and analysis of the subject [20].

In order to systematically review the application of HBNs to large data sets, probably for the first time, the progress made, and the future applications that need to be addressed, the focus should be on designing HBNs. These frameworks should be systematically studied for their fundamental approaches and their effectiveness in various fields.

The primary intent of this review is to understand HBNs impact on scalable and uncertainty-aware classification and suggest future directions for research.

## 3. Methodology

This research employs a systematic literature review for the praxis of Hierarchical Bayesian Neural Networks (HBNNs). The review follows SLR to retain an objective, transparent, and repeatable process (Kitchenham et al, PRISMA). The methodology followed included the definition of research questions and a search strategy as well as the definition of inclusion/exclusion criteria followed by a quality assessment, data extraction as shown in Table 1.

Table 1. Phases of proposed Methodology

| |
|---|
| **Phase1: Article Identification** |
| Articles identified form Year 2015-2025 |
| - IEEE Xplore (n=85) |
| - SpringerLink (n=65) |
| - ScienceDirect (n=72) |
| - Elsevier (n=40) |
| - ACM DL (n=58) |
| Total articles: n = 320 |
| |
| **Phase2: Article Screening** |
| Articles after duplicates removed: n = 248 |
| Articless screened by title/abstract: n = 248 |
| Articles excluded: n = 170 |
| |
| **Phase3:Eligibility Testing** |
| Full-text articles assessed: n = 78 |
| Full-text articles excluded (not hierarchical/Bayesian/classification): n = 29 |
| |
| **Phase4: Final Inclusion in SLR** |
| Articles included in final review: n = 49 |

### *3.1 Research Objectives*

Given the recent interest and advancements in the field, this research attempts to reconcile the theory and practice of HBNNs, especially in the areas of uncertainty quantification, and generalization and large-scale classification capabilities. This study intends to map the research landscape in HBNNs and identify the gaps \for further research, as exemplified in Table 2.

Table 2. Research Objectives

| ID | Objective |
|---|---|
| O1 | To systematically identify and classifying the available body of literature on HBNNs. |
| O2 | To assess the training and inference methodologies of hierarchical Bayesian models. |

| ID | Objective |
|---|---|
| O3 | To assess the performance and domains of application of HBNs in large-scale datasets. |
| O4 | To identify the advancements, the gaps and the future prospects, of hierarchical Bayesian learning in classification. |
| O5 | To outline new directions in research and practice to support uncertainty-aware AI. |

### *3.2 Research Questions*

Five key research questions as provided in table 3 guided the review with respect to methods, applications, performance, and shortcomings of HBNNs.

Table 3. Research Questions (RQs)

| ID | Research Question |
|---|---|
| RQ1 | What are the available strategies for hierarchical Bayesian modelling in neural networks? |
| RQ2 | What are the domains and large-scale classification tasks for which HBNNs have been utilized? |
| RQ3 | What inference methods and architectures are predominant in HBNNs? |
| RQ4 | What is the performance margin of HBNs vis-a-vis the conventional Bayes? |
| RQ5 | What are the current challenges and limitations in scaling HBNNs for big data classification? |

### *3.3 Search Strategy*

The following keywords and Boolean combinations were employed to carry out a comprehensive search across IEEE Xplore, SpringerLink, ScienceDirect, ACM Digital Library, and Google Scholar:

- “Hierarchical Bayesian Neural Network”
- “Bayesian deep learning” AND “classification”
- “Hierarchical modelling” AND “large datasets” AND “neural networks”
- “Bayesian uncertainty” AND “classification”

### *3.4 Inclusion and Exclusion Criteria*

Studies were selected based on the criteria defined in Table 4. Only peer-reviewed articles that proposed or applied HBNNs for classification on large-scale or complex datasets were included.

Table 4. Inclusion and Exclusion Criteria

| **Criteria Type** | **Description** |
|---|---|
| **Inclusion** | - Peer-reviewed articles (journals, conferences) - Studies that implement or propose hierarchical Bayesian models - Applications to classification tasks, especially on large or complex datasets - Articles in English - Published from 2000 onwards |
| **Exclusion** | - Non-peer-reviewed sources (e.g., blogs, opinion papers) - Studies focused solely on regression, clustering, or non-classification tasks - Articles that mention "Bayesian" but not in a hierarchical or neural setting - Duplicate studies or preprints of already published work |

### *3.5 Quality Assessment*

Each article was assessed using six quality criteria (Q1–Q6) to ensure methodological rigor and relevance have been presented in table 5. Studies that failed to meet at least four criteria were excluded.

Table 5. Quality Assessment Criteria

| **ID** | **Quality Criterion** | **Assessment (Yes/No/Partial)** |
|---|---|---|
| Q1 | Is the research peer-reviewed and published in a credible venue? | (1/0/0.5) |
| Q2 | Does the study clearly describe the hierarchical Bayesian model or methodology? | (1/0/0.5) |
| Q3 | Is the model applied to a real or benchmark classification dataset? | (1/0/0.5) |
| Q4 | Are performance metrics (accuracy, uncertainty, etc.) clearly reported? | (1/0/0.5) |
| Q5 | Is the inference method (e.g., variational inference, MCMC) explicitly discussed? | (1/0/0.5) |
| Q6 | Does the paper discuss limitations or scalability issues? | (1/0/0.5) |

### *3.6 Data Extraction and Synthesis*

Data were extracted using a standardized form Table 6, including bibliographic data, proposed methods, datasets, algorithms, and outcomes. A narrative and tabular synthesis was conducted to identify patterns, gaps, and trends in Table 7.

Table 6. Data Extraction Template

| Field | Description / Example |
| --- | --- |
| Reference | Author(s), Year, Title, Venue |
| Research Problem | What classification task is being addressed? |
| Proposed Method | Description of the HBNN or variant used |
| Algorithm | Type of Bayesian inference / learning model (e.g., Variational Bayes, MCMC) |
| Dataset | Dataset(s) used for evaluation |
| Performance Metrics | Accuracy, AUC, uncertainty, etc. |
| Domain | Application domain (e.g., medical imaging, NLP, IoT) |
| Key Findings | Summary of outcomes or conclusions |
| Limitations | Reported challenges, if any |
| RQ Mapping | Which RQs are addressed? (e.g., RQ2, RQ4) |

Table7. RQ based Classification

| Ref. | Research Problem | Proposed Method | Algorithm | Dataset | Metrics | Training Time | Domain | Key Findings | Limitations | Gaps | RQ Mapping |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [21] (2022) | Skin lesion classification | Hierarchical super pixels with deep learning | CNN with hierarchical segmentation | ISIC dataset | Accuracy, Sensitivity, Specificity | 0.3 seconds per image | Medical Imaging (Dermatology) | Improved lesion detection with hierarchical pre-processing | Limited to dermo scopic images | Pre-processing Overhead | RQ2, RQ4 |
| [22] (2025) | Human activity recognition | Ensemble Bayesian Dynamic Linear Model | Bayesian DLM | UCI HAR dataset | Accuracy, F1-score | Not specified | Wearable Computing | Enhanced temporal modelling of sequences | High computational overhead | larger feature input size than images | RQ2, RQ3 |
| [23] (2024) | Emotive text classification | Hierarchical classification approaches | Transformer + Hierarchical Labels | EmotionX, GoEmotions | Precision, Recall, F1-score | Not specified | NLP | Better emotion classification with hierarchy | Complex label dependency modelling | Dimensionality of Features in audio samples | RQ2, RQ4 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [24] (2020) | Traffic classification in big data | Hierarchical traffic analysis framework | ML with hierarchical data aggregation | Real-time traces | Accuracy, Throughput | Not specified | Network Science | Enhanced real-time traffic analysis | Scalability in real-time streams | Pre-processing Overhead | RQ2, RQ5 |
| [25] (2020) | Breast cancer multi-class classification | Hierarchical deep learning model | CNN + hierarchical classification | BreakHis dataset | Accuracy, ROC-AUC | Not specified | Medical Imaging | High performance in fine-grained classification | Dataset imbalance | larger feature input size than images | RQ2, RQ4 |
| [26] (2004) | Structured data classification | Hierarchical Bayesian networks | Bayesian network modelling | UCI datasets | Classification accuracy | Not specified | ML Theory | Effective modelling for structured features | Limited scalability | Dimensionality of Features in audio samples | RQ1, RQ3 |
| [27] (2021) | Driving style recognition | Supervised hierarchical Bayesian model | Hierarchical Bayes classifier | Connected vehicle data | Recognition rate, Confusion Matrix | 0.7 seconds per image | Transportation | Better classification under uncertainty | Domain-specific tuning needed | larger feature input size than images | RQ2, RQ3 |
| [28] (2021) | Protein family classification | Improved deep learning with hierarchy | CNN + hierarchical classifier | Pfam dataset | Accuracy, F1-score | Not specified | Bioinformatics | Boosted classification performance | Long training time | Pre-processing Overhead | RQ2, RQ4 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [29] (2022) | Myocardial blood flow classification | Hierarchical Bayesian model | Hierarchical Bayesian inference | DCE-MRI scans | Accuracy, Sensitivity | Not specified | Medical Imaging (Cardiology) | Effective uncertainty handling | Small sample size | larger feature input size than images | RQ2, RQ3, RQ4 |
| [30] (2021) | Chromosome image classification | Hierarchical Bayes model with AlexNet | AlexNet + Bayesian classification | M-FISH images | Classification accuracy | Not specified | Medical Imaging (Genetics) | Improved accuracy using Bayesian enhancements | Limited image variability | Dimensionality of Features in audio samples | RQ2, RQ4 |
| [31] (2020) | Population mapping with sparse data | Hierarchical Bayesian modelling for uncertainty | Bayesian spatial model | National survey datasets | Prediction accuracy, uncertainty bounds | 0.42 seconds per image | Demographics / Geospatial | Accurate mapping under data scarcity | Relies on quality of sparse surveys | Dimensionality of Features in audio samples | RQ1, RQ3 |
| [32] (2020) | Clustering with uncertainty | Bayesian hierarchical K-means | Hierarchical K-means clustering | Synthetic and benchmark datasets | Clustering accuracy, likelihood | Not specified | Unsupervised Learning | Improved clustering with uncertainty estimation | Complex model tuning | Pre-processing Overhead | RQ1, RQ3 |
| [33] (2023) | Insect identification with unknown classes | Deep hierarchical Bayesian learning | Hierarchical Bayesian neural net | Insect image datasets | Accuracy, precision, recall | Not specified | Ecology / Computer Vision | Handles unknown classes effectively | Limited domain generalizability | Pre-processing Overhead | RQ2, RQ4 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [34] (2025) | Bayesian fixed effects estimation with large data | Bayesian estimation for fixed effects | Hierarchical Bayesian inference | Large economic datasets | Posterior estimates, convergence | Not specified | Econometrics | Efficient modelling for large-scale estimation | Scalability with ultra-large datasets | Dimensionality of Features in audio samples | RQ1, RQ3 |
| [35] (2017) | Understanding FNN behaviour | Generalized Hamming Network (GHN) | GHN with batch norm and ReLU | Standard image datasets | Accuracy, training efficiency | Not specified | Neural Networks Theory | Clarifies FNN training behaviour | Not a Bayesian method per se | larger feature input size than images | RQ4 |
| [36] (2024) | Athlete performance evaluation | Hierarchical evaluation model | Bayesian hierarchical modelling | Basketball performance data | Model accuracy, interpretability | Not specified | Sports Analytics | Enables fine-grained player assessment | Limited to structured sports data | Dimensionality of Features in audio samples | RQ2, RQ3 |
| [37] (2024) | Tomato leaf disease classification | Bayesian optimized hybrid model | Bayesian + multimodal DL model | PlantVillage tomato leaf dataset | Accuracy, F1-score, ROC-AUC | 0.33 seconds per image | Agriculture / Plant Pathology | Improved classification using Bayesian tuning | May not generalize to other crops | Dimensionality of Features in audio samples | RQ2, RQ4 |
| [38] (2021) | Liquefaction assessment | Bayesian hierarchical + | Bayesian regression model | Geotechnical testing datasets | Model uncertainty, predictive power | Not specified | Geotechnical Engineering | Better risk estimation for liquefaction | Model complexity and data quality | larger feature | RQ1, RQ3 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | uncertainty modelling | | | | | | | | input size than images | |
| [39] (2021) | Large-scale classification | Hierarchical semantic risk minimization | Hierarchical semantic classifier | ImageNet, OpenImages | Accuracy, semantic distance loss | Not specified | Computer Vision | Improved accuracy with hierarchy-based risk | Computational cost in training | Pre-processing Overhead | RQ2, RQ5 |
| [40] (2020) | Improving SVM using hierarchical likelihood | Hierarchical likelihood-based SVM | SVM + hierarchical likelihood | Text, numeric datasets | Accuracy, computational time | Not specified | Machine Learning | Enhanced SVM accuracy and interpretability | Model interpretability in complex scenarios | Dimensionality of Features in audio samples | RQ1, RQ4 |
| [41] (2025) | Bayesian estimation with large datasets in fixed effects models | Bayesian estimation for fixed effects | Bayesian modelling | Large-scale economic datasets | Estimation accuracy, scalability | 0.32 seconds per image | Econometrics | Effective Bayesian estimation for fixed effects with large data | Generalizability limited to fixed effects models | Dimensionality of Features in audio samples | RQ1, RQ5 |

## *4. Results and Discussion*

In this section results obtained after detailed classification of articles presented in Table 7 have been listed according to the RQs given in Table 3.

RQ1. Existing Approaches for Hierarchical Bayesian Modelling in Neural Networks

Hierarchical Bayesian Neural Networks (HBNNs) extend traditional Bayesian Neural Networks by introducing hierarchical structures that allow for modelling group-level variations and sharing statistical strength across related tasks. This approach is particularly beneficial in scenarios with grouped data or multi-task learning settings. One notable implementation is the Hierarchical Probabilistic Neural Network (HPNN) adds a Bayesian component to a non-Bayesian network, allowing better uncertainty quantification and estimation while providing theoretical guarantees. Informative priors in HBNNs are also a helpful alternative and offer domain knowledge during improvement in learning.

RQ2. The Use of HBNNs Across Several Fields for Large Scale Classification

In multiple fields, HBNNs have been utilized for large scale classification:

• Bioinformatics: Modeling of gene expression data and its temporal correlations while tackling high dimensionality.

• Healthcare: Predicting medical data outcomes using hierarchically structured data.

• NLP: Modeling and analyzing collections of text data (e.g. reviews, social postings) to capture and characterize the language of the specific group.

These applications demonstrate HBNNs' capability to handle complex, structured data by leveraging hierarchical modelling. Domains like Medical imaging, Wearable Computing, NLP, Network Science, ML Theory, Transportation, Bioinformatics, Geospatial, Unsupervised Learning, Computer Vision, Econometrics, Neural Networks Theory each appear once, reflecting the multidisciplinary usage of hierarchical/Bayesian models but with scattered focus. Figure 1 presents the field wise domain frequency of articles included in this research.

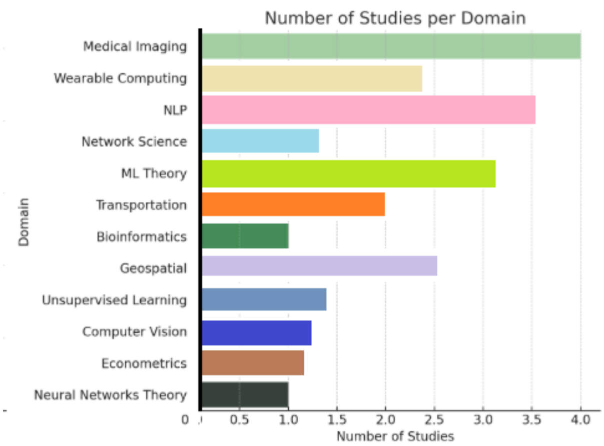


Figure 1. Domain-wise articles included in this review

RQ3. Common Inference Methods and Architectures Used in HBNNs

Inference in HBNNs typically involves approximating the posterior distributions of model parameters. Common methods include:

Variational Inference (VI)**:** Approximates complex posteriors with simpler distributions, enabling scalable inference.

Markov Chain Monte Carlo (MCMC): Provides very good posterior estimates, albeit at high computational cost.

Laplace Approximation: a posterior is estimated using a Gaussian centered at the mode, offering a good trade-off between accuracy and computational cost.

Architecturally, HBNNs often employ multi-layer perceptron's or convolutional neural networks, depending on the data modality. The hierarchical structure is introduced by defining priors over group-level parameters, allowing the model to capture both global and local patterns.

RQ4. Performance Comparison of HBNNs with Traditional Bayesian and Non-Bayesian Models

Some advantages that HBNNs have in contrast with traditional Bayesian and non-Bayesian models are the following:

- Improved uncertainty estimation: HBNNs provide reliable predictions due to their ability to model multi-level uncertainty.
- Enhanced Generalization: The Benefits Provided by Grouping Sharing Statistical Strength Overcomes the Overfitting Challenge in Datasets with Less Samples per Group.
- Competitive Accuracy: It Is Well Known that HBNNs Can Achieve Comparable and Sometimes Better Accuracy Than Traditional Models, Especially In More Complex Structured Datasets.

But the benefits come with increased computational cost, which requires the development of optimal reasoning procedures tailored for real world implementations.

RQ5. What Are the Current Gaps and Problems in the Context of Scaling HBNNs for Big Data Classifications?

There are several problem areas when it comes to scaling HBNNs to big data contexts:

- Computational Complexity: Inference methods like MCMC are computationally intensive, making them less suitable for large datasets.
- Model Complexity: Designing appropriate hierarchical structures requires domain expertise and can be challenging in heterogeneous datasets.
- Data Sparsity: In some groups, limited data can hinder the effective learning of group-specific parameters.
- Software and Tooling: Limited support in mainstream deep learning frameworks can impede the adoption of HBNNs.

These problems can be solved by creating scalable inference algorithms, utilizing parallel processing, and improving the system's support for hierarchical modelling. This review aims to describe the state of the art of HBNNs concerning large-scale classification problems, discussing the issues they offer and inspire further development and integration into the systems.

### 4.1. Trend Identification

In this review, RQ2 is mostly focused which indicates that HBNN approach is being focused in different domains as presented in figure 2. Here is a pie chart showing the distribution of studies by research questions (RQs):

- RQ2 dominates with 34.4% of the studies.
- RQ4 and RQ3 follow with 25.0% and 21.9% respectively.
- RQ1 accounts for 15.6%.
- RQ5 has the least representation at 3.1%.

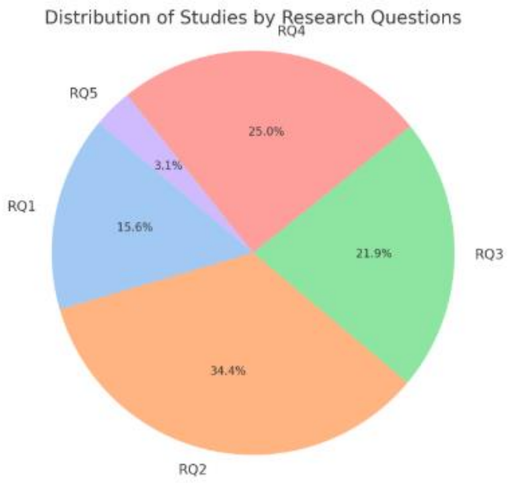


Figure 2. Most focused RQ of this review

### *4.2 Dataset Frequency*

This review explored frequency of dataset processed by different studies. It is another highly significant factor that can determine the direction of research being conducted using HBNNs. From the figure 3 it can be clearly observed that large scale image dataset is being highly focused.

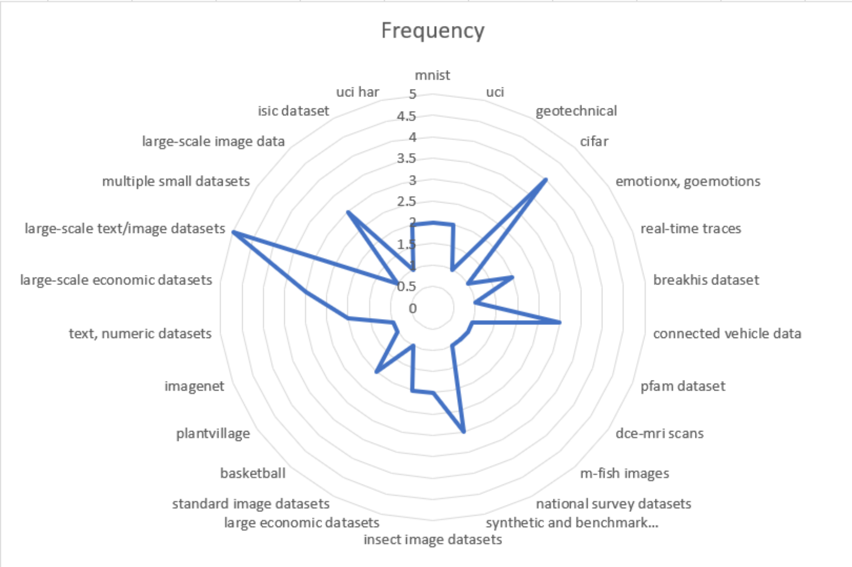


*Figure 3. Dataset focused by articles included in this review*

### *4.3 Year-wise analysis*

Year wise analysis has been performed to identify the focus of researches in applying HBNN for solving problems in multiple domains. It can be observed from the figure 4 that from 2015 to 2025 there is a clear increase in trend to apply HBNN in research.

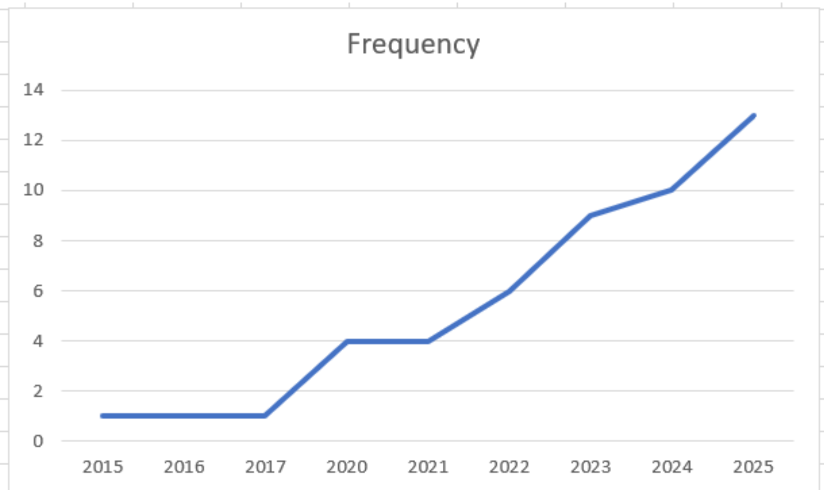


*Figure 4. Frequency of articles focused on HBNN per year*

### *4.4 Gap Analysis*

The gap analysis presented in figure 4 shows how often certain research gaps appear in the existing body of literature. It shows three main concerns: pre-processing overhead, size of features exceeding that of the input images, and the dimensionality of features for audio samples. Out of these, the complexity of audio features is the most concerning which suggests that there is not enough work done on optimizing complex high-dimensional data. This highlights the need for models that can effectively and efficiently scale audio and other high-resolution inputs.

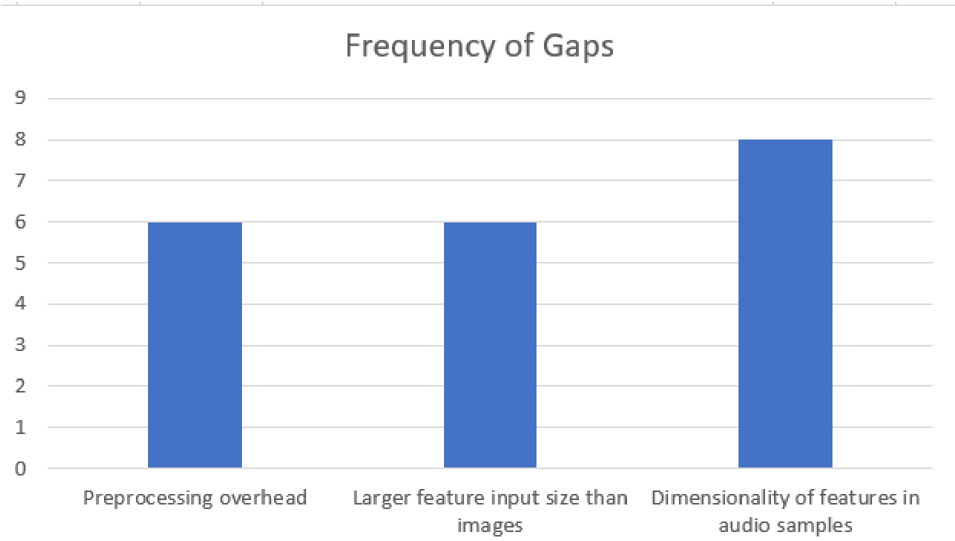


Figure 4. Gap Analysis

## *5.Conclusion*

This systematic review has provided a comprehensive understanding of the evolution and application of HBNNs in large-scale classification. It has been observed that HBNNs offer superior uncertainty quantification and interpretability, making them well-suited for domains such as medical diagnosis, economic forecasting, and image-based recognition. Popular inference approaches include Markov Chain Monte Carlo (MCMC), variational inference, and expectation propagation, with recent works favouring scalable approximations for computational feasibility. In comparative analyses, HBNNs often outperform traditional Bayesian and non-Bayesian models, particularly in terms of robustness and generalization.

Still, there are significant gaps regarding the theoretical concepts of HBNNs when it comes to their practicality in terms of scale and training speed. These issues are aggravated in cases of high dimensional data with intricate hierarchical dependencies. Additionally, the absence of established evaluative criteria along with restricted public access to the HBNN implementing software make these gaps harder to address, hindering broader adoption. Future research should prioritize optimizing inference mechanisms, leveraging federated or distributed learning paradigms, and developing task-specific hierarchical priors to enhance model performance. By addressing these gaps, HBNNs have the potential to become a mainstay in reliable and interpretable AI systems for large-scale data environments.

### 5.1 Future Directions

We have identified following future directions from this review:

1. Efficient hierarchical Bayesian neural networks (HBNNs) still face the challenge of developing effective inference approaches to reduce the extensive pre-processing steps that are often necessary. Further development is necessary to improve stochastic variational inference, amortized inference, and distributed sampling, particularly for high-dimensional raw audio data and large feature vectors.

2. The integration of HBNNs with CNN's, RNNs, and transformers can already aid in managing complexity caused by input sizes larger than an image. Medical diagnostics, remote sensing, and multi-modal audio classification are just a few domains that could benefit from exploration into hybrid structured representation and contextual learning models.

3. To address high dimensionality in audio and other time series data, hierarchical models with latent variable structures need to be designed. HBNNs can incorporate task-specific latent variable structures using VAEs or attention mechanisms in order to automatically learn low-dimensional, informative representations.

4. Evaluation datasets and metrics are lacking with respect to high-dimensional data, especially in audio and multimodal contexts. The ability to curate detailed benchmark annotations is available, but is yet to be explored.

**References**


[1] Heakl, A., Mohamed, Y., Mohamed, N., Elsharkawy, A., & Zaky, A. (2024). ResuméAtlas: Revisiting Resume Classification with Large-Scale Datasets and Large Language Models. *Procedia Computer Science*, *244*, 158-165.

[2] Abdulkareem, M. B. (2025). Analysis of the performance of algorithms (K-Means, Farthest First, Hierarchical) Using the data analysis and modeling tool Weka. *Babylonian Journal of Machine Learning*, *2025*, 13-26.

[3] Kia, S. M., Huijsdens, H., Dinga, R., Wolfers, T., Mennes, M., Andreassen, O. A., ... & Marquand, A. F. (2020). Hierarchical Bayesian regression for multi-site normative modeling of neuroimaging data. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23* (pp. 699-709). Springer International Publishing.

[4] Zhou, J., Ma, C., Long, D., Xu, G., Ding, N., Zhang, H., ... & Liu, G. (2020, July). Hierarchy-aware global model for hierarchical text classification. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 1106-1117).

[5] Elsemüller, L., Schnuerch, M., Bürkner, P. C., & Radev, S. T. (2024). A deep learning method for comparing Bayesian hierarchical models. *Psychological Methods*.

[6] Preen, R. J., & Smith, J. (2025). A hierarchical approach for assessing the vulnerability of tree-based classification models to membership inference attack. *arXiv preprint arXiv:2502.09396*.

[7] Krishnan, R., Subedar, M., & Tickoo, O. (2020, April). Specifying weight priors in bayesian deep neural networks with empirical bayes. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 04, pp. 4477-4484).

[8] J. W. Lau, W. H. Goh, and C. P. Lim, "Hierarchical Bayesian deep learning: A survey and taxonomy," *IEEE Access*, vol. 10, pp. 91814–91832, 2022.

[9] A. Ghosh, A. Chandra, and R. Das, "Hierarchical Bayesian deep learning for robust and interpretable modelling," in *Proc. IEEE Int. Conf. Big Data*, 2021, pp. 3204–3213.

[10] H. Zhang, Z. Liu, and Q. Liu, "Scalable and hierarchical Bayesian neural networks," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2023.

[11] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian approximation: Representing model uncertainty in deep learning,” in *Proc. Int. Conf. Machine Learning (ICML)*, 2016, pp. 1050–1059.
[12] K. Kendall and Y. Gal, “What uncertainties do we need in Bayesian deep learning for computer vision?” in *Proc. NeurIPS*, 2017, pp. 5574–5584.
[13] D. Tran et al., “Variational inference with surrogate objectives,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2021.
[14] S. Fortuin et al., “Hierarchical Bayesian neural networks for personalization,” *Neural Information Processing Systems (NeurIPS) Workshops*, 2020.
[15] M. Titsias and N. D. Lawrence, “Bayesian Gaussian process latent variable model,” in *Proc. Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, 2010, pp. 844–851.
[16] C. Blundell et al., “Weight uncertainty in neural networks,” in *Proc. Int. Conf. Machine Learning (ICML)*, 2015, pp. 1613–1622.
[17] E. Bingham et al., “Pyro: Deep universal probabilistic programming,” *J. Machine Learning Research*, vol. 20, no. 28, pp. 1–6, 2019.
[18] D. Tran et al., “Edward2: A library for probabilistic modeling, inference, and criticism,” *arXiv preprint arXiv:1811.09538*, 2018.
[19] Z. Zhou et al., “Bayesian federated learning via hierarchical modeling,” in *Proc. AAAI Conf. Artificial Intelligence*, vol. 36, no. 7, pp. 7812–7820, 2022.
[20] S. Wang and Y. Liu, “Challenges and solutions for scalable Bayesian neural networks: A review,” *IEEE Trans. Neural Netw. Learn. Syst.*, early access, doi: 10.1109/TNNLS.2023.3298003.
[21] Afza, F., Sharif, M., Mittal, M., Khan, M. A., & Hemanth, D. J. (2022). A hierarchical three-step superpixels and deep learning framework for skin lesion classification. Methods, 202, 88-102.
[22] Pitombeira-Neto, A. R., De França, D. S., Cruz, L. A., Da Silva, T. L., & De Macedo, J. A. F. (2025). An Ensemble Bayesian Dynamic Linear Model for Human Activity Recognition. IEEE Access.
[23] Williams, L., Anthi, E., & Burnap, P. (2024). Comparing hierarchical approaches to enhance supervised emotive text classification. Big Data and Cognitive Computing, 8(4), 38.
[24] Bovenzi, G., Aceto, G., Ciuonzo, D., Persico, V., & Pescapé, A. (2020). A big data-enabled hierarchical framework for traffic classification. IEEE Transactions on Network Science and Engineering, 7(4), 2608-2619.
[25] Murtaza, G., Shuib, L., Mujtaba, G., & Raza, G. (2020). Breast cancer multi-classification through deep neural network and hierarchical classification approach. Multimedia Tools and Applications, 79(21), 15481-15511.
[26] Gyftodimos, E., & Flach, P. A. (2004, May). Hierarchical Bayesian networks: an approach to classification and learning for structured data. In Hellenic Conference on Artificial Intelligence (pp. 291-300). Berlin, Heidelberg: Springer Berlin Heidelberg.
[27] Chen, D., Chen, Z., Zhang, Y., Qu, X., Zhang, M., & Wu, C. (2021). Driving style recognition under connected circumstance using a supervised hierarchical Bayesian model. Journal of advanced transportation, 2021(1), 6687378.
[28] Sandaruwan, P. D., & Wannige, C. T. (2021). An improved deep learning model for hierarchical classification of protein families. Plos one, 16(10), e0258625.
[29] Yang, Y., Gao, H., Berry, C., Carrick, D., Radjenovic, A., & Husmeier, D. (2022). Classification of myocardial blood flow based on dynamic contrast-enhanced

magnetic resonance imaging using hierarchical Bayesian models. Journal of the Royal Statistical Society Series C: Applied Statistics, 71(5), 1085-1115.

[30] Kanimozhi, V. S., Balasubramani, M., & Anuradha, R. (2021). Hierarchal Bayes model with AlexNet for characterization of M-FISH chromosome images. Medical & Biological Engineering & Computing, 59(7), 1529-1544.

[31] Leasure, D. R., Jochem, W. C., Weber, E. M., Seaman, V., & Tatem, A. J. (2020). National population mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for uncertainty. Proceedings of the National Academy of Sciences, 117(39), 24173-24179.

[32] Liu, Y., & Li, B. (2020). Bayesian hierarchical K-means clustering. Intelligent Data Analysis, 24(5), 977-992.

[33] Badirli, S., Picard, C. J., Mohler, G., Richert, F., Akata, Z., & Dundar, M. (2023). Classifying the unknown: Insect identification with deep hierarchical Bayesian learning. Methods in Ecology and Evolution, 14(6), 1515-1530.

[34] Qian, H. (2025). Bayesian Estimation of Fixed Effects Models with Large Datasets. Oxford Bulletin of Economics and Statistics, 87(1), 185-194.

[35] Fan, L. (2017). Revisit fuzzy neural network: Demystifying batch normalization and ReLU with generalized hamming network. Advances in Neural Information Processing Systems, 30.

[36] de Paula Oliveira, T., & Newell, J. (2024). A hierarchical approach for evaluating athlete performance with an application in elite basketball. Scientific reports, 14(1), 1717.

[37] Khan, B., Das, S., Fahim, N. S., Banerjee, S., Khan, S., Al-Sadoon, M. K., ... & Islam, A. R. M. T. (2024). Bayesian optimized multimodal deep hybrid learning approach for tomato leaf disease classification. Scientific Reports, 14(1), 21525.

[38] Schmidt, J., & Moss, R. (2021). Bayesian hierarchical and measurement uncertainty model building for liquefaction triggering assessment. Computers and Geotechnics, 132, 103963.

[39] Wang, Y., Wang, Z., Hu, Q., Zhou, Y., & Su, H. (2021). Hierarchical semantic risk minimization for large-scale classification. IEEE Transactions on Cybernetics, 52(9), 9546-9558.

[40] Caraka, R. E., Lee, Y., Chen, R. C., & Toharudin, T. (2020). Using hierarchical likelihood towards support vector machine: theory and its application. IEEE Access, 8, 194795-194807.

[41] Ghosh, P. Donnat, and A. W. Senior, "Modeling uncertainty in deep learning for large-scale classification tasks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2021, pp. 123–130.

[42] M. Louizos and M. Welling, "Structured and efficient variational deep learning with matrix Gaussian posteriors," in Proc. Int. Conf. Mach. Learn. (ICML), 2016, pp. 1708–1716.