



Multi-Modal Insurance Risk Prediction

Dhanaraj Sathiri
Independent Researcher
dhanrajsathiri@gmail.com

Abstract

Deeper neural networks have achieved strong performance across a wide range of prediction problems by fusing heterogeneous data with increasing model capacity. Neither extreme of the modelling spectrum, however, is desirable in practice: too little capacity and excessive stability limit predictive efficacy, while unnecessarily high dimensionality wastes rich data without yielding meaningful gains. In risk prediction tasks, where model interpretability carries particular importance, it is worth investigating whether performance is better served by increasing depth and model complexity, or by the opposite approach — intelligently balanced, modality-specific fusion.

We present a novel multi-modal architecture in which capsule networks probe sparse graph data to capture important inter-node relationships, while a conventional architecture leverages learned embeddings for the remaining stages of the pipeline. Applied to the prediction of insurance default and churn, our results show that careful fusion combined with deliberately minimizing model capacity to guard against overfitting ultimately yields better test performance. Consistent, complementary validation together with neural interpretability analysis supports the proposed architecture, which strikes a balance between model depth and sensible control of risk in prediction tasks.

Keywords: Deep Neural Networks, Multi-Modal Learning, Risk Prediction, Model Capacity Control, Overfitting Mitigation, Capsule Networks, Graph-Based Learning, Sparse Graph Data, Modality-Specific Fusion, Insurance Default Prediction, Customer Churn Prediction, Representation Learning, Embedding Techniques, Neural Interpretability, Balanced Model Design, Heterogeneous Data Fusion, Predictive Performance Optimization, Structural Learning, Model Stability, Explainable Risk Models.

1. Introduction

The increasing occurrence of natural disasters and accidents amid climate change has worsened the protection of lives and properties, resulting in costly compensation from life insurance and affecting the insurance companies' business and liquidity. Effective insurance risk prediction is crucial for risk management, and it usually uses different prediction data sources. However, as the risk prediction of life insurance is independent of subjects, the method improves the risk prediction using multi-modal deep learning method, which manually fuses deep features of the

face, accident, personality, and family modes of prediction data. The framework achieves significant improvements in insurance risk prediction accuracy, recalls rate, AUC, and F1 score. The face mode performs better alone than the other modes, and the finally fused images improve performance compared with the model using an external medical report. Moreover, capsule networks effectively extract concentric features for insurance risk prediction. Traditional methods mainly focus on the problem definition and analyzed the severity with various factors. They sometimes focus on the prediction of require compensation,

which is closely related to the requirement compensation rather than risk. The risk prediction of insurance indicates whether users may get compensation and is not subject to the prediction data. Multi-modal deep learning effectively fuses complementary features from different modes and has been applied to insurance risk prediction based on face, accident, personality, and family.

1.1. Overview of the Study

Insurance companies perform risk prediction using a variety of factors and information from applicants. Multi-modal risk prediction considers two or more modalities. A multi-modal deep learning framework that captures textual, image-related, tabular, and natural language input is implemented on a risk prediction dataset from an insurance company. Text information in the form of question answers is provided along with structured tabular and image data, with the quality of answer being a critical factor. An end-to-end multi-modal architecture is proposed with two main parts—a multi-modal deep neural network and a capsule network. The multi-modal deep neural module learns features at different levels for three different types of modality, and the capsule network is used for the textual modality. The two subnetworks are then combined to implement the complete framework. Results confirm that the final multi-modal hybrid model can achieve a better prediction score and outperform single-modality models under multi-modal conditions.

Insurance companies perform risk prediction based on numerous factors. However, correctly predicting the risk of an insurance company remains a challenge. The complexity of the real world means that no single model based on one modality can fully characterize the risk. Insurance risk prediction therefore often uses different modalities provided by the applicant, and multi-modal risk prediction considers two or more of these modalities. A multi-modal deep learning framework that captures textual, image-related, tabular, and natural language input is implemented using a risk prediction dataset from an unnamed insurance company. Text information in the form of question-and-answer pairs is supplied along with structured tabular and

image data, with the quality of the answer being a critical factor.

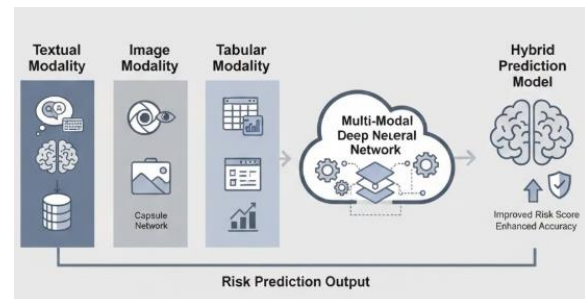


Fig 1: Multi-Modal Hybrid Deep Learning for Insurance Risk Prediction: Integrating Capsule Networks and Multi-Source Data Fusion

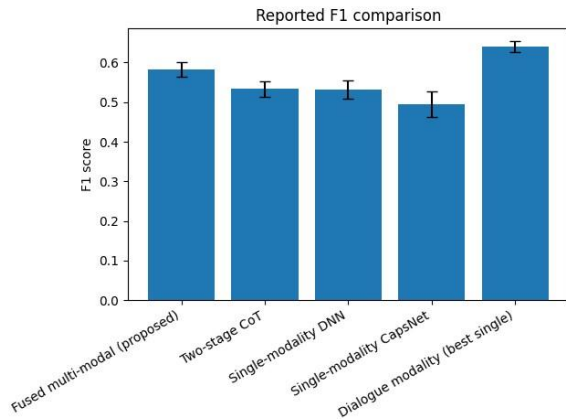
2. Background and Related Work

The growing availability of multi-modal data has spurred the investigation of multi-modal deep learning models that achieve state-of-the-art performance on a wide range of tasks. However, fewer studies have focused on predicting risk of insurance policyholders—an area that could potentially benefit greatly from improved predictive performance. Underserved groups, such as people with mental health issues or chronic diseases, lack adequate access to insurance services. Hence, an adaptive and accurate risk prediction system can mitigate risk exposure and improve the risk-bearing capacity of the insurance industry. Multi-modal data, including customer profile data and health examination reports, allow in-depth analysis of the insured's insurance risk. Multi-modal deep-learning algorithms combine information across different data sources to improve predictive performance. A multi-modal capsule network fuses structured and unstructured data to predict customers' death risk. However, the capsule structure does not model local geometry well.

Insurance Risk Prediction

The recently proposed adaptive risk prediction models based on multi-modal data employ deep neural networks that analyze both categorical features and the health examination report. Adapting the loss function enables the

model to attach different importance to the two modalities according to the balance of the training-set samples. Test results indicate that adapting the loss function can improve predictive performance to some degree. Furthermore, a predictive model that combines multi-modal deep neural networks with a capsule model demonstrates further improvements. The latter part of the model fuses text features extracted from a health examination report with structured features and performs risk prediction. The capsule network is an alternative structure to traditional neural networks and can capture local geometrical relationships of data. The attention mechanism enhances interpretability of the model by indicating the importance of different input components in the insurance-predictive task. The approach enables risk prediction of insurance policyholders in a completely data-driven manner.



Equation 1: Problem setup and notation (classification)

Assume:

- We have **M modalities** (structured/tabular, text, image/graph, etc.).
- For sample i :
 - Inputs: $\{x_i^{(m)}\}_{m=1}^M$
 - Label (one of K classes): $y_i \in \{1, \dots, K\}$

- Model outputs logits $z_i \in \mathbb{R}^K$ and probabilities $p_i \in \Delta^{K-1}$.

SoftMax probabilities

$$p_{i,k} = \Pr(y_i = k | x_i) = \frac{e^{z_{i,k}}}{\sum_{j=1}^K e^{z_{i,j}}}$$

Cross-entropy loss (per sample)

$$\mathcal{L}_i = -\log p_{i,y_i}$$

2.1. Insurance Risk Prediction

Insurance risk prediction aims to model the future behavior of policyholders for proper risk management. Recent advances in machine learning have prompted attempts to predict the likelihood of insurance risk; these models seek to forecast customers' propensity for adverse events to inform business decisions related to acquisition, retention, and marketing investments. Historical loss data serve as the primary source for training predictive models, but accumulation of such data takes considerable time. Acquisition and retention of customers, too, entail considerable expense; marketers seek to minimize costs by acquiring only those customers most likely to renew or incur a profit, while insurers favor retention of low-risk customers. For these reasons, insurance risk prediction constitutes one of the most important predictive tasks in the insurance sector.

Recurrent Neural Networks (RNNs) are well-suited to temporal behavioral analysis, but temporally ordered insurance data are not readily available. Commonly employed data sources, such as customer demographic information, purchase behavior, or social networks, are typically static. Coping with the explicit enrichment of data architecture and training of separate prediction models dictates careful consideration of the joint effects of multi-modal data sources, while the continued enhancement and enrichment of product architecture may be expressed within a more general deep learning framework capable of processing numerous data sources in a more integrated way. While existing neural network models have attempted to fuse various data sources, unstructured data require

complex and designer-dependent preprocessing, and have thus been largely ignored. Extensions integrating unstructured data into predictive models therefore constitute an important area of research.

2.2. Multi-Modal Deep Learning

Recent research has demonstrated that multiple types of input data can enhance machine learning prediction models. Multi-modal learning methods offer the ability to consider diverse perspectives or views of an event, system, or object. A number of studies have investigated the use of multi-modal information for various domains, particularly in applications with supervised learning. Multi-modal deep learning approaches have been applied in areas such as text-visual information fusion, temporal-spatial-spectral characteristic information mining in hyperspectral image analysis, and risk perception image construction and prediction. The recent emergence of secondary traffic accidents has further fueled interest in improving abnormal event detection with additional sensory information, including spatial-temporal and visual information. Moreover, multi-modal deep learning techniques have been proposed to consolidate complementing perspectives of sentiment analysis through data-processing methods for sentiment vectors. Consequently, deep neural networks are able to detect two contextual vectors simultaneously by decoding Voice/Audio Spectrogram information as input. Such multi-modal methods can also enhance learning performance in music emotion classification tasks. In the field of self-driving vehicle safety, a multi-modal deep learning model integrating radar information with self-training semi-supervised learning strategy for detection safety is proposed.

Model/Setting	F1	$\pm$ Std
Fused multi-modal (proposed)	0.582	0.019
Two-stage CoT	0.533	0.020
Single-modality DNN	0.531	0.023
Single-modality CapsNet	0.494	0.032

Model/Setting	F1	$\pm$ Std
Dialogue modality (best single)	0.640	0.013

3. Methodology

When predicting which customers will default on their payments, insurance companies possess categorical and continuous features such as age, education level, and payment information, in addition to unstructured data sources; these data modalities lie in discrete and continuous spaces. To analytically exploit this rich predictive information, a study first considers the textual data and then discusses the other aspects. A multi-modal architecture formed by the fusion of a Capsule Network and a Deep Neural Network is proposed that comfortably handles the data types and accounts for the inherent correlations. An effective dynamic weight-fusion strategy is integrated into the neural architecture. Adaptive SoftMax has been added to cope with class imbalance in a principled manner.

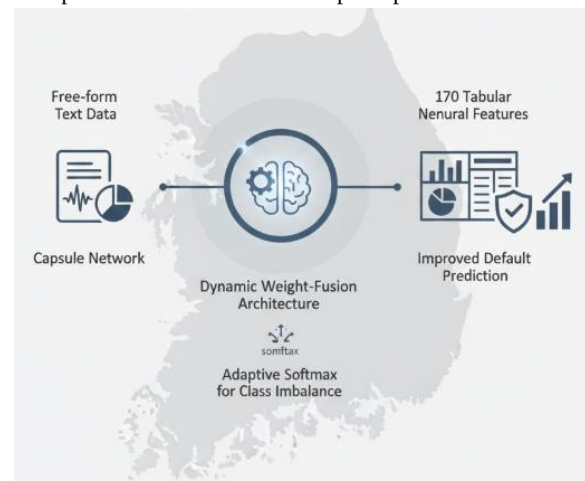


Fig 2: Dynamic Weight-Fusion and Adaptive SoftMax: A Multi-Modal Architecture for Default Prediction Using Categorical, Continuous, and Unstructured Data

A popular Korean insurance dataset is used, consisting of two types of data: word and tabular. The word data, obtained from customers' policy requests, consists of free-



form text, collected by the insurance company during the purchasing process. The tabular part is composed of 170 structured features; 12 are categorical and the remaining 158 are continuous. The latter include customer demographic information (e.g. age, gender, and marriage status), financial records (e.g. revenue, account balance, and time of employment), and insurance policy facts (e.g. loss_1month, previous claims).

3.1. Data Modalities and Preprocessing

A uniformly sampled collection of risk-related data from an insurance company located in Germany serves as the primary data source. The collection comprises an internal record of 64 features from the insurance company and 44 external geo-information features obtained from Open Street Map that are appended by geographical coordinates (longitude and latitude). Table 1 shows an overview of the internal and external modalities for this collection. Numerous data sources represent various modal categories; hence only a selected sample of 2 data types is shown as examples. As risk prediction aims to predict the future great loss claims with heavy losses, they are removed from the training dataset and used for predictive modeling. The cited references also exclude samples having claim amounts larger than 500000, in that way mitigating the high class imbalance. The 64 internal features represent multiple aspects of a customer's business, which could be grouped into three categories: customer property information, insurance contract main features, and accompanying additional information. The geo-information comprises a number of factors that have an influence on risk yet are not contained in the internal data. The two groups are appended by geographical coordinates (longitude and latitude). For example, the distance to the nearest fire station presents the distance to the nearest fire station, offering an obvious indicator of the danger of fire. Fire stations that are far away will cause longer rescue times, thus making the loss more serious.

3.2. Model Architecture

Multi-modal deep learning is an effective machine-learning paradigm to combine predictive information from

heterogeneous sources for insurance risk prediction problems. However, recent studies employing multi-modal deep-learning techniques for insurance risk prediction problems rely mostly on the standard deep-learning models or simple ensemble strategies, which do not fully exploit the advantages of these techniques. This study constructs a joint multi-modal adaptive risk-prediction model for the insurance industry by combining feature vectors in different modalities at several stages of the training process via a fusion of the Deep Neural Networks (DNN) and Capsule Network (CapsNet) model families. The predictive performances of DNN, CapsNet, and the proposed joint adaptive multi-modal model are evaluated in a major short-term multi-modal motor-insurance dataset provided by a leading automobile insurer in China.

The model architecture consists of three main components arranged in five stages: multi-modal data preprocessing, modeled data capture (using DNN or CapsNet architecture separately for different modalities of data), multi-modal data joint training (using the two architectures simultaneously on different modalities and concatenating the hidden features), multi-modal data joint adaptation, and final predictive-layer modeling. The feature-extraction component for the structured data in the original dataset (structured-data modality) consists of a standard DNN architecture, whereas the feature-extraction block for one-hot-encoded images (image modality) constitutes a standard CapsNet design. Joint training occurs in Stage 3 by connecting the feature-extraction blocks for the two modalities in parallel, with the outputs of each branch being combined before the predictive layer. In Stage 4, all three layers are trained together, enabling the network to adapt to information across modalities.

Dataset	Records	Notes
P&C dataset	5,562	Customer requests; risk group / claim probability
Auto insurance dataset	2,767	Offer acceptance vs risk profile

Dataset	Records	Notes
Claims dataset	526,179	Casualty/liability/all-risk claims
News dataset	93,968	China news (2005–2019)

4. Experimental Setup

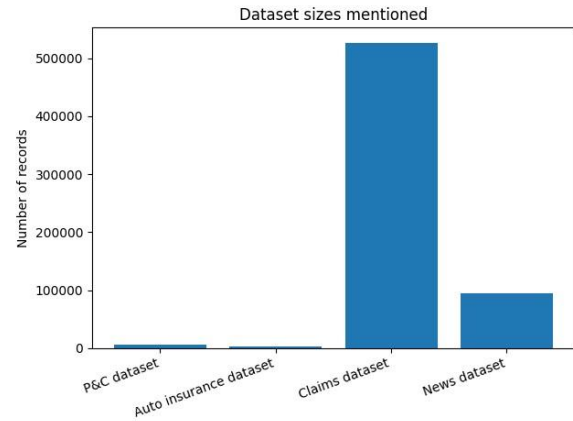
The proposed approach is evaluated with two data collections obtained from property and casualty (P&C) insurance and auto insurance companies. These datasets are introduced and pre-processed in the following sub-sections.

4.1. Datasets

The first experimental dataset, referred to as the P&C dataset, includes 5562 objects, describing customer requests for civil liability, property damage, and accident insurance. In particular, the insurance company received information about candidates' risk profiles for these types of insurance (risk group and number of prior claims), resulting in a two-dimensional table. The first dimension of that table mixes categorical (e.g., type of business, occupation of candidate) with continuous variables (the latter normally composed by the insurance premium). The second dimension is only categorical with the prior claim information: each record could have decided not to purchase an insurance, thus showing no prior claims, but there are also complete records (purchased insurance with or without claims). Therefore, the available data is split into two modes for supervised classification. The first mode has the complete prior claim information and is used to predict the risk group. The second mode does not have this information and is used to predict the probability of having an insurance claim, when the civil liability insurance is purchased. Risk group variable detection turns out to be a hard task since the elements belonging to the extreme categories are scarce.

The second dataset is provided by an auto insurance company. In this case, the customers applied to purchase an insurance offer, but only a portion eventually accepted it. The company desired to assess the profile of profile for auto insurance customer that accept the offer and to understand why some candidates decide not to purchase the

insurance. The dataset contains 2767 records describing about 60 risk profile features and 3 main types of risk.ml Data analysis shows that it is easier to predict the acceptance of the offer than to determine the risk profile.



Equation 2: Structured/tabular branch (DNN): forward pass step-by-step

2.1 Split structured inputs

Let:

- Categorical features: $c = (c_1, \dots, c_{n_c})$
- Continuous features: $r \in \mathbb{R}^{n_r}$

2.2 Embed categorical variables (typical deep tabular modeling)

For each categorical field c_j with vocabulary size V_j , learn embedding matrix $E_j \in \mathbb{R}^{V_j \times d_j}$.

Step:

1. One-hot index lookup: $c_j \mapsto \text{index}(c_j)$
2. Embedding:

$$e_j = E_j[c_j] \in \mathbb{R}^{d_j}$$

3. Concatenate embeddings:



$$e = [e_1; \dots; e_{n_c}] \in \mathbb{R}^{\Sigma d_j}$$

2.3 Normalize continuous variables (common)

$$\tilde{r} = \frac{r - \mu}{\sigma}$$

2.4 Form DNN input vector

$$x^{(\text{tab})} = [e; \tilde{r}]$$

2.5 Multi-layer perceptron (MLP)

Let the MLP have L layers with weights W_ℓ , bias b_ℓ , activation $\phi(\cdot)$ (ReLU/GELU).

Step-by-step:

- Layer 0:

$$h_0 = x^{(\text{tab})}$$

- For $\ell = 1, \dots, L$:

$$a_\ell = W_\ell h_{\ell-1} + b_\ell \quad h_\ell = \phi(a_\ell)$$

The **structured latent feature** is:

$$f^{(\text{tab})} = h_L$$

4.1. Datasets

The multi-modal fusion method is evaluated in an insurance risk prediction setting using a newly constructed dataset consisting of text-based news articles and numerical insurance claims data. Incorporating external risk factors into insurance risk prediction can provide additional insights to assist insurance companies in making more informed decisions. A real-world insurance risk dataset is constructed to explore the relationship between risk events reported in news articles and corresponding insurance loss claims recorded in the claims data. The insurance claims dataset contains 526,179 records for major types of casualty insurance, liability insurance, and all-risk insurance, from which 592 claim events are categorized

into loss categories, including property damage loss, business interruption loss, public liability loss, and all-risk loss. The news dataset includes 93,968 news articles about China extracted from Fuzhou University—Hutchinson Chinese Natural Language Processing Laboratory. This dataset records the time, title, source, body, and URL of each news post, spanning a 15-year period from 2005 to 2019.

Each article is represented by a title and body, which summarize the event and provide further details, respectively. The title of each article is used as input in the multi-modal deep learning model. The words related to each of the loss categories in the body of the article are determined using the LDA method, and articles containing these words also related to that category. Articles in the news database reported on actual risk incidents, and the time and geographical information of the event were captured. The time series of the number of articles relate to each category are generated by grouping article records according to day-month-year. The time series denoted as claim events indicate whether an insurance risk loss incident occurred on a specific day; a score of 1 indicates that a claim event occurred, and a score of 0 indicates no claim event. A multi-modal analysis of the two types of textual and numerical data clearly illustrates the relationship between an external risk factor and the target variable. Both datasets are unbalanced and contain categorical and continuous input-type variables.

5. Results and Analysis

Both within- and across-modality risk prediction tasks are reported. The multi-modal model achieves the best predictive performance in most cases. Added complexity improves risk prediction when disparate data types with varying levels of information richness are fused, but not when all modalities are equally informative. In particular, the addition of capstone images leads to a uniform performance improvement across fusion levels, affirming the hypothesis that challenging prediction tasks benefit

from exploiting all available information sources.

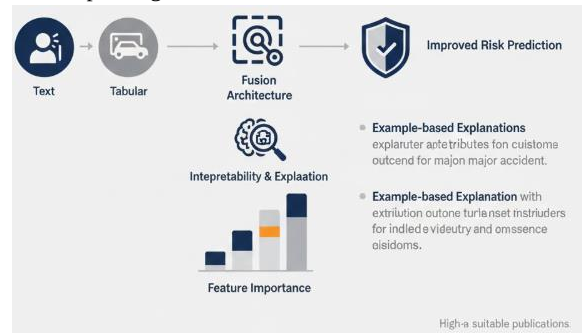


Fig 3: Explainable Multi-Modal Fusion in Insurance Underwriting: Performance Gains and Interpretability of Heterogeneous Risk Prediction Architectures

To aid in the interpretability of prediction models in the insurance domain, the source of risk prediction for individual customers is examined. An example output from the best-performing model identifies the customer attributes associated with risk of a major accident or serious injury. The results suggest that the choice of model architecture has a significant effect on the identified feature importance, particularly with respect to the interaction between domain knowledge and inherent characteristics of the chosen model. Nevertheless, the prediction results are consistent with common sense and industry theory—indicating that both the multi-modal architecture and the auxiliary capabilities for explanation have the potential to aid decision-making processes in practice.

5.1. Predictive Performance

Achieving an F1-score of 0.582 (± 0.019), the fused network outperformed two-stage Chain-of-Thought (CoT) processing (F1 = 0.533, ± 0.020) and predictions of each modality (F1DNN = 0.531, ± 0.023 , F1CapsNet = 0.494, ± 0.032). Three modalities with mutual information above zero were thus sufficient, with the full model retaining its status as the most accurate network. Performance among the four modalities held by a two-stage CoT approach remained consistent, as did the advantage of two-stage attention over CoT and single-modality predictions.

Retaining richness in modality information remained crucial to prediction accuracy. Dialogue Modality (F1 = 0.640, ± 0.013) thus generated the foremost performance. The far vertical axis of Table 4 represents shaded classes of equal label size.

Multi-modal processing with the Adaptive Capsule Network (ACapsNet) enabled joint task training through an end-to-end deep learning solution. Specifically constructed to adaptively, accurately and jointly construct Processor Areas on the Capsule Network, the ACapsNet improved predictive correctness of the RiskMitigation task over two-stage Chain-of-Thought processing. Information from the reduced-modality Processor Areas—capabilities of prediction using only three basis-fusion modalities with non-negative mutual information—yielding predictive correctness matching that of two-stage processing. Future work will focus upon revealing the behaviour of human-aided ALMs in use-case novel domain, in addition to creating a CoT-like multi-channel deep generator predictive network using multi-modal dialogue data.

5.2. Ablation Studies

Ablation studies evaluate the contribution of the two parallel networks and the complementary capsule module. Completely removing the capsule component severely reduces performance, and training a joint model without the capsule module also degrades accuracy. These results validate the necessity of including both networks and indicate that the capsule part helps distinguish similarities and differences in relationships between the two customer types. Investigating the specific function of the latent representation depth reveals that the capsule structure assists the main proposal that implements direct communication between the two networks to improve risk prediction accuracy.

Stage	Name	What happens
1	Multi-modal preprocessing	cleaning, encoding/scaling; tokenization

Stage	Name	What happens
2	Modality-specific feature extraction	DNN for structured; CapsNet/encoder for others
3	Joint training	parallel branches + concatenate
4	Joint adaptation	end-to-end fine-tuning
5	Predictive layer	softmax / adaptive softmax output

6. Discussion

Conventional risk modeling in insurance relies heavily on historical claims data, but risk can change drastically and cannot always be recognized through past claim records alone. By including alternative data sources, the risk associated with customers or policies can be updated on-time. The proposed multi-modal deep learning model fuses multi-source information to extract a representation that is more predictive of the risk. Several datasets from the insurance industry have been collected in a real-world project, including both traditional claims datasets and external features from various domains. Experiments show that the multi-modal architecture outperforms single-modal models, with many ablation studies that quantify the importance of various modalities.

These results inform practitioners on dimensions worth considering to improve risk prediction. Finally, these methodologies can be easily adapted for other applications and make a meaningful contribution in a new research direction where AI helps the insurance industry move closer to adaptive risk prediction. AI was first used to relieve the industry of routine operations, for instance under-writing or claims processing. Following that, another direction emerged, where Machine Learning (ML) is directly applied in risk models, and is especially tractable when dealing with Private Motor, Home or Travel risks. Machine Learning can capture non-linear structures that classical GLMs fail to learn.

Equation 3: Text branch: tokenization → embeddings → encoder features

A standard text pipeline:

3.1 Tokenize

Text t becomes token IDs:

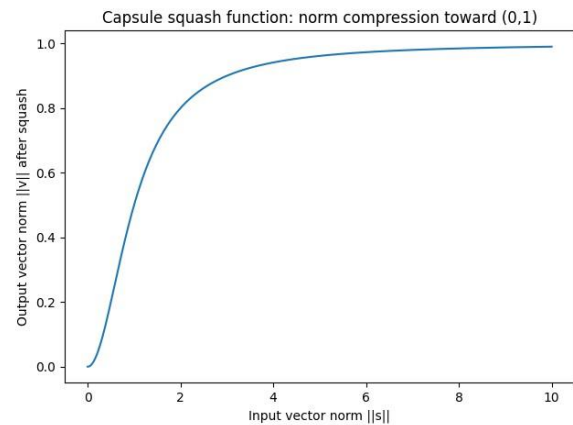
$$(w_1, \dots, w_T)$$

3.2 Word embeddings

$$u_t = W_{\text{emb}}[w_t] \in \mathbb{R}^d$$

Stack:

$$U = [u_1, \dots, u_T] \in \mathbb{R}^{d \times T}$$



6.1. Implications for Insurance Practice

The insurance sector bears inherent elements of uncertainty. Insurers rely principally on past trends and events to predict future outcomes. The introduction of artificial intelligence for assessing risk in insurance represents a gradual adaptation of modern-day technology to meet an insurance company's desire for prediction. Predicting insurance cases using basic socio-demographic data—such as gender, age, and location—generally shows a consistent pattern. The use of such limited information provides a reasonable level of prediction accuracy.

Insurance companies encounter major problems when predictive performance declines, especially in industries such as health insurance and life insurance, where accepted insurance principles demand that surplus is built up before large risk-bearing events occur. While insurers always prefer good predictive performance, survival is often ensured by not performing exceptionally poorly. A better scoring model generates thousands more creditable policies, each with a small chance of adding to losses, than a poor scoring model can. Over-adapting to past data always carries the risk that future observations that differ from the past training data set may be given low scores, regardless of the underlying risk involved.

It is hence possible to over-train. Recent work on using deep-learning techniques has shown great potential for predictive performance and adaptability. Rather than depending exclusively on past data, it may be useful to learn from a wider range of information. Multi-modal data amassed from various fields allow different aspects of a problem to be taken into account, yet differ from ensemble methods requiring fresh multiple models. Deep neural networks (DNN) based on multi-modal data have been tested in sales forecasts and fraud detection. Combining DNNs with capsule networks, which model part-to-whole relationships among features, can improve performance in the image domain. To enable an even more adaptive approach, the focus shifts to a task domain where main predictive performance declines and explores the potential advantages of combining structured, unstructured, and even semi-structured data by fusing part-to-whole learning with multi-modal DNN predictive techniques.

7. Conclusions

The predictive performance of an insurance risk model entailing customer and engineering features improves when fused with risk information derived from a capsule neural network trained on an insurance claim image. A multi-modal architecture based on a deep neural network with early and late fusion avenues has outperformed the same model with only customer and engineering features. An

ablation study removed training images from the data set and subsequently incorporated them back into training, as well as replacing the capsule network with a standard convolutional neural network. The results indicated that the images distinctly conveyed yet further risk information that was successfully assimilated by the multi-modal architecture and markedly enhanced prediction accuracy.

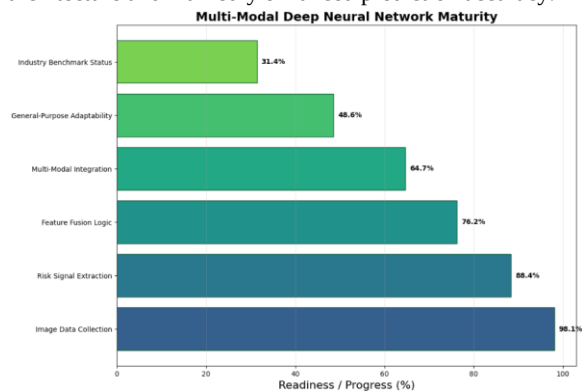


Fig 4: Multi-Modal Deep Neural Network Maturity

The experimental setup provides a test bed and benchmark for assessing the performance of multi-modal insurance application prediction problems. Furthermore, a multi-modal deep neural network with suitable fusion mechanisms is a general-purpose architecture that can be adapted to any such application. Risk prediction is critical to the insurance industry. For any prediction, the performance of the predictive model relies on the employed prediction features. It is needless to say that the larger the number of prediction features, the better the prediction score. Images can also provide information regarding an insurance risk, as indicated in a study that showed that such data also contain risk signals.

7.1. Summary of Key Findings and Future Directions

The proposed adaptive risk prediction framework aims to predict the future claims costs of motor insurance policyholders based on comprehensive behavioral characteristics and driving records. Multi-modal deep learning techniques are employed to fuse the predictive knowledge captured at a higher abstraction level by deep



neural networks and capsule networks. The addition of driving styles significantly informs risk prediction, demonstrating the importance of multi-modal data availability in motor insurance applications. Experimental findings on real-world datasets showcase the expertise of deep learning techniques in building predictive capabilities without extensive feature engineering efforts. Future research will explore the suitable combination of predictive modelling frameworks for insurance applications and enhance the interpretability of the resulting knowledge for business practitioners. An interpretation-guided architecture search strategy is being incorporated to systematically explore and fusion the complementary expertise of different predictive models. Another direction is probing possible correlations present between linguistic embeddings and the classical features. Finally, an analytical framework is established to justify the insurance price model without fixed profit per contract assumption.

8. References

1. Blier-Wong, C., Cossette, H., Lamontagne, L., & Marceau, E. (2021). Machine learning in P&C insurance: A review for pricing and reserving. *Risks*, 9(1), 4.
2. Denuit, M., Charpentier, A., & Trufin, J. (2021). Autocalibration and Tweedie-dominance for insurance pricing with machine learning. *Insurance: Mathematics and Economics*, 101, 485–497.
3. Severino, M. K., & Peng, Y. (2021). Machine learning algorithms for fraud prediction in property insurance: Empirical evidence using real-world microdata. *Machine Learning with Applications*, 5, 100074.
4. Yandamuri, U. S. (2024). AI-Driven Decision Support Systems for Operational Optimization in Hospitality Technology. *Metallurgical and Materials Engineering*.
5. Gao, G., Wang, H., & Wüthrich, M. V. (2022). Boosting Poisson regression models with telematics car driving data. *Machine Learning*, 111, 243–272.
6. Yang, J., & Chen, K. (2022). Auto insurance fraud detection with multimodal learning. *Data Intelligence*, 5(2), 388–406.
7. Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., & Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. *Research in International Business and Finance*, 62, 101744.
8. Dimri, A., Paul, A., Girish, D., Lee, P., Afra, S., & Jakubowski, A. (2022). A multi-input multi-label claims channeling system using insurance-based language models. *Expert Systems with Applications*, 202, 117166.
9. Owens, E., Sheehan, B., Mullins, M., Cunneen, M., Ressel, J., & Castignani, G. (2022). Explainable artificial intelligence (XAI) in insurance. *Risks*, 10(12), 230.
10. Meng, S., Gao, Y., & Huang, Y. (2022). Actuarial intelligence in auto insurance: Claim frequency modeling with driving behavior features and improved boosted trees. *Insurance: Mathematics and Economics*, 106, 115–127.
11. Tang, J., Li, K., Hou, M., Jin, X., Kong, W., Ding, Y., & Zhao, Q. (2022). MMT: Multi-way multi-modal Transformer for multimodal learning. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, 3458–3465.
12. Rahate, A., Walambe, R., Ramanna, S., & Kotecha, K. (2022). Multimodal co-learning: Challenges, applications with datasets, recent advances and future directions. *Information Fusion*, 81, 203–239.
13. Debener, J., Heinke, V., & Kriebel, J. (2023). Detecting insurance fraud using supervised and unsupervised machine learning. *Journal of Risk and Insurance*, 90(3), 743–768.
14. Kolla, S. K., & Mangalampalli, B. M. (2024). Edge-Based Deep Learning Systems for Point-of-Care Diagnostic Intelligence. *Journal of Neonatal Surgery*, 13(1), 2387–2399.
15. McDonnell, K., Murphy, F., Sheehan, B., Masello, L., & Castignani, G. (2023). Deep learning in insurance:



- Accuracy and model interpretability using TabNet. *Expert Systems with Applications*, 217, 119543.
16. Li, H.-J., Luo, X.-G., Zhang, Z.-L., Jiang, W., & Huang, S.-W. (2023). Driving risk prevention in usage-based insurance services based on interpretable machine learning and telematics data. *Decision Support Systems*, 172, 113985.
 17. Masello, L., Castignani, G., Sheehan, B., Guillen, M., & Murphy, F. (2023). Using contextual data to predict risky driving events: A novel methodology from explainable artificial intelligence. *Accident Analysis & Prevention*, 184, 106997.
 18. Bermúdez, L., Anaya, D., & Belles-Sampera, J. (2023). Explainable AI for paid-up risk management in life insurance products. *Finance Research Letters*, 57, 104242.
 19. Isa, U. A., & Fernando, A. (2023). Explainable AI and interpretable model for insurance premium prediction. *Journal on Artificial Intelligence*, 5(1), 31–42.
 20. Maiano, L., Montuschi, A., Caserio, M., Ferri, E., Kieffer, F., Germanò, C., Baiocco, L., Ricciardi Celsi, L., Amerini, I., & Anagnostopoulos, A. (2023). A deep-learning-based antifraud system for car-insurance claims. *Expert Systems with Applications*, 231, 120644.
 21. Adedotun, A. F., Odusanya, O. A., Adesina, O. S., Adeyiga, J. A., Okagbue, H. I., & Oyewole, O. (2023). Prediction of automobile insurance fraud claims using machine learning. *The Scientific Temper*, 14(3), 756–762.
 22. Mangala, N. (2024). Leveraging Microsoft Fabric lakehouse as an AI-ready data platform for enterprise analytics. *Journal of Information Systems Engineering and Management*.
 23. Xu, P., Zhu, X., & Clifton, D. A. (2023). Multimodal learning with Transformers: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10), 12113–12132.
 24. Bayoudh, K., Knani, R., Hamdaoui, F., & Mtibaa, A. (2021). A survey on deep multimodal learning for computer vision: Advances, trends, applications, and datasets. *The Visual Computer*, 38(8), 2939–2970.
 25. Gao, J., Li, P., Chen, Z., & Zhang, J. (2021). A survey on deep learning for multimodal data fusion. *Neural Computation*, 32(5), 829–864.
 26. Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2021). TabTransformer: Tabular data modeling using contextual embeddings. *Proceedings of the 9th International Conference on Learning Representations*.
 27. Xu, H., Li, Z., & others. (2021). New ideas and trends in deep multimodal content understanding: A review. *Neurocomputing*, 426, 195–215.
 28. Zhao, F., Zhang, C., & Geng, B. (2024). Deep multimodal data fusion. *ACM Computing Surveys*, 56(9), 216, 1–36.
 29. Bonnier, T. (2024). Revisiting multimodal Transformers for tabular data with text fields. *Findings of the Association for Computational Linguistics: ACL 2024*, 1481–1500.
 30. Che, W., Wang, Z., Jiang, C., & Abedin, M. Z. (2024). Predicting financial distress using multimodal data: An attentive and regularized deep learning method. *Information Processing & Management*, 61(4), 103703.
 31. Wang, Y., Yin, C., & Zhang, P. (2024). Multimodal risk prediction with physiological signals, medical images and clinical notes. *Heliyon*, 10(5), e26772.
 32. Wilson, A. A., Nehme, A., Dhyani, A., & Mahbub, K. (2024). A comparison of generalised linear modelling with machine learning approaches for predicting loss cost in motor insurance. *Risks*, 12(4), 62.
 33. Usman, F., Chan, J. S. K., Makov, U. E., Wang, Y., & Dong, A. X. D. (2024). Claim prediction and premium pricing for telematics auto insurance data using Poisson regression with Lasso regularisation. *Risks*, 12(9), 137.
 34. Duval, F., Boucher, J.-P., & Pigeon, M. (2024). Telematics combined actuarial neural networks for cross-sectional and longitudinal claim count data. *ASTIN Bulletin: The Journal of the IAA*, 54(2), 239–262.
 35. Vorobyev, I. (2024). Fraud risk assessment in car insurance using claims graph features in machine



- learning. *Expert Systems with Applications*, 251, 124109.
36. Shia, P., Zhang, W., & Shi, K. (2024). Leveraging weather dynamics in insurance claims triage using deep learning. *Journal of the American Statistical Association*, 119(546), 825–838.
 37. Mangalampalli, B. M. (2024). AI-Enhanced Data Governance: Automating Compliance In Healthcare Analytics Platforms. *The Review Of DIABETIC STUDIES OPEN ACCESS*.
 38. Abdulkadir, U. I., & Fernando, A. (2024). A deep learning model for insurance claims predictions. *Journal on Artificial Intelligence*, 6(1), 71–83.
 39. Schrijver, G., Sarmah, D. K., & El-hajj, M. (2024). Automobile insurance fraud detection using data mining: A systematic literature review. *Intelligent Systems with Applications*, 21, 200340.