# Outer Bound for Rate-Reliability-Equivocation Region of Compound Wiretap Channel with Informed Terminals

Mariam Haroutunian[1] [0000−0002−9262−4173]

Institute for Informatics and Automation Problems,
National Academy of Sciences of Armenia, Yerevan, Armenia
armar@sci.am

**Abstract.** The goal in designing communication systems in the presence of a wiretapper is to ensure that the message remains confidential between the transmitter and the intended receiver while minimizing the information available to the eavesdropper. Here we investigate the compound wiretap channel model, which is the extension of the wiretap channel, when the channels to the legitimate receiver and to the wiretapper depends on the number of possible states. Various cases can be considered, when these states are known or unknown to legitimate terminals.
We investigate the E-capacity-equivocation region which is the closure of the set of all achievable rate-reliability and equivocation pairs, where the rate–reliability function represents the optimal dependence of rate on the error probability exponent (reliability). Here the outer bound of this region is constructed in the case, when the states of the main channel are known to the legitimate terminals. A similar result for the case with unknown states was published previously.

**Keywords:** Compound Wiretap Channel · Channel Capacity · Rate-Reliability-Equivocation Region.

## 1 Introduction

A **wiretap channel**, in the context of communication systems and information theory, refers to a communication channel that is tapped or intercepted by a third party with the intention of eavesdropping on the communication. This concept is often used in the study of secure communication and cryptography.

In a wiretap channel model, there are typically three parties involved [1]:

*Transmitter*: This is the source that is trying to send a message to a legitimate receiver.

*Receiver*: The intended recipient of the message.

*Eavesdropper*: A third party that is intercepting or tapping into the communication channel in an attempt to gain unauthorized access to the message.

The transmitter wishes to send a message $m$ to the receiver while keeping it as secret as possible from the eavesdropper. The information-theoretic investigation of generalized model of wiretap channel can be found in [2] - [6]. The wiretap

# Application of Identification Codes to the Two-Party Privacy-Preserving Record Linkage (PPRL) *

Yanling Chen[0000−0003−1603−9121]

Volkswagen Infotainment GmbH, Bochum, Germany
yanling.chen@volkswagen-infotainment.com

**Abstract.** In this paper, we apply the identification codes to the problem of two-party privacy-preserving record linkage (PPRL). In particular, we emphasize the advantage of our approach on the performance analysis, especially on the privacy analysis, over the classical hash-based approaches. Note for the PPRL, linkage quality is typically evaluated experimentally, whilst for privacy, there is so far no commonly accepted privacy measures available that allow an objective evaluation. Our approach of identification code provides an objective evaluation on both linkage quality and privacy based on parameters of identification codes.

**Keywords:** Record Linkage · Identification code · Privacy

## 1 Introduction

Privacy preserving record linkage (PPRL) addresses the problem of linking records that represent the same individuals across several datasets without revealing sensitive information of the individuals [4,5]. So far a variety of linkage protocols have been proposed. See a short list that includes but not limited to [6,9,11,12].

In general, proposals to PPRL can be classified into those that require a third party for performing the linkage and those that do not. The former are known as 'three-party protocols' and the latter as 'two-party protocols'. In three-party protocols, a (trusted) third party (which we call the 'linkage unit') is involved in conducting the linkage, while in two-party protocols only the two database owners participate in the PPRL process. In this paper, we put our focus on the two-party protocols.

Generally, two-party protocols start by the two database owners agreeing upon and exchanging any required information such as parameter settings, preprocessing methods, encoding or encryption methods, and any secret keys that

* The main part of the work was conducted as the author was with University of Duisburg-Essen, and was supported by the German Research Foundation under the research grant DFG 407023611.

# An Automated Approach to Collecting and Labeling Time Series Data for Event Detection Using Elastic Node Hardware

Tianheng Ling, Islam Mansour, Chao Qian, and Gregor Schiele

Intelligent Embedded Systems Lab, University of Duisburg-Essen,
47057, Duisburg, Germany
{tianheng.ling, chao.qian, gregor.schiele}@uni-due.de
islam.mansour@stud.uni-due.de

**Abstract.** Recent advancements in IoT technologies have underscored the importance of using sensor data to understand environmental contexts effectively. This paper introduces a novel embedded system designed to autonomously label sensor data directly on IoT devices, thereby enhancing the efficiency of data collection methods. We present an integrated hardware and software solution equipped with specialized labeling sensors that streamline the capture and labeling of diverse types of sensor data. By implementing local processing with lightweight labeling methods, our system minimizes the need for extensive data transmission and reduces dependence on external resources. Experimental validation with collected data and a Convolutional Neural Network (CNN) model achieved a high classification accuracy of up to 91.67%, as confirmed through 4-fold cross-validation. These results demonstrate the system's robust capability to collect audio and vibration data with correct labels.

**Keywords:** Event Detection · Time Series · Sensor Data Collection · Automated Labeling · Embedded Systems · CNN · Integrated Hardware System

## 1 Introduction

Event detection has become a popular topic in pervasive computing [1], enabling intelligent systems to interpret environmental contexts and adapt configurations within various spaces, for example, offices or kitchens [2, 3]. Traditional IoT methods often utilize multiple types of indirect sensor data, such as audio and vibrations [4], which are processed through Deep Learning (DL) models for event recognition.

Sufficiently labeled datasets are necessary to train DL models effectively [5]. Typically, data streams are segmented and annotated with labels [6]. One common approach to collecting these datasets involves transmitting sensor data to the cloud [7], where labeling algorithms are applied [8], or storing the data streams for subsequent manual labeling by human workers [9]. Both methods, however, introduce significant delays and dependencies on external resources.

# Towards Training DNNs with Quantized Parameters

Leo Buron[1,2][0009−0001−8939−4784], Lukas Einhaus[1,2][0000−0002−6102−7077], Andreas Erbslöh[1,2][0000−0001−6702−892X], and Gregor Schiele[1,2][0000−0003−4266−4828]

[1] University of Duisburg-Essen, LAB for Intelligent Embedded Systems, Forsthausweg 2, 47057 Duisburg, Germany firstname.lastname@uni-due.de https://www.uni-due.com/es/

[2] paluno - The Ruhr Institute for Software Technology, University of Duisburg-Essen, Essen, Germany https://paluno.uni-due.de/

**Abstract.** For embedded devices, memory and computational cost are limited. However, the training of neural networks is computation and memory expensive. To reduce memory consumption quantization schemes are applied on the parameters by a lot of related work. The introduced artifacts can be reduced by applying stochastic rounding instead of round-half-to-even to the quantization scheme. While a lot of related work is using stochastic rounding for inference and gradient computation, none have shown the benefit of it for the parameter update only. We use a fixed point quantization scheme and propose to quantize parameters at model initialization. When training, we use full-resolution gradient computation and apply stochastic quantized updates. The quantization of the parameters and the parameter update lead to reduced memory consumption. In addition, the inference is mostly quantized, speeding up computations. For training, the inputs of each layer are stored for gradient computations. Due to quantized inference those inputs are already quantized reducing the needed memory further. We explore a model for different fixed point configurations on the FASHION-MNIST dataset. Generally, stochastic rounding parameter updates match or beat the top-5 accuracy of QAT. We are able to achieve 0.876 top-5 accuracy while reducing the memory to 0.93. When using stochastic rounding as well for inference, we can achieve 0.737 top-5 accuracy for a memory reduction of 0.88.

**Keywords:** deep learning · quantized parameters · edge training · embedded devices · memory reduction · computational cost · stochastic round-ing · inference · fixed point quantization

## 1 Introduction

With the recent success of deep neural networks, more challenging tasks are solved with more complex and bigger neural networks. The increase in network parameters makes the training challenging even on GPU clusters, where memory is often the bottleneck to fit these models on GPUs [5]. This is a more

# Ensembling Machine Learning Models for Malware Detection

Patricio Galdames[1,2][0000−0003−3051−2413],
Claudio Gutierrez-Soto[3][0000−0002−7704−6141], and
Marco A. Palomino[4][0000−0001−7850−416X]

[1] Universidad San Sebastián, Concepción, Chile, patricio.galdames@uss.cl
[2] Whitecliffe College, Christchurch, New Zealand, 20230676@mywhitecliffe.com
[3] Universidad del Bío-Bío, Concepción, Chile, cogutier@ubiobio.cl
[4] University of Aberdeen, Scotland, UK, marco.palomino@abdn.ac.uk

**Abstract.** Malware poses serious problems for individuals and businesses worldwide. No matter how much we do to prevent it, attackers continue to find ways to challenge preventive strategies. Thus, we propose a new ensemble learning methodology to enhance malware detection. Our proposal improves predictive accuracy and generalization capabilities by using support vector machines, random forests, and neural networks, all of them trained with publicly available databases and statically extracted features. We have also explored meta-learner construction approaches, such as stacking and weighted voting, to optimally integrate our base detectors, and we have evaluated our ensemble comprehensively, using F1 score, Matthew's correlation coefficient, and informedness metrics, among others. The results demonstrate the potential of ensemble learning for a robust malware defense and highlight the value of informative features, adaptive retraining, and comprehensive evaluation. Clearly, our approach offers better performance than its individual components, particularly in minimizing false positives and negatives.

**Keywords:** Malware Detection, Ensemble Learning, Machine Learning, Static Analysis, Meta-Learners

## 1 Introduction

The fourth industrial revolution, marked by the growing integration of sensors into products and processes, has led to a massive increase in data generation. Organizations must deploy advanced technologies to process these data effectively. However, the value of organizational data has also attracted specialized hackers, known as Advanced Persistent Threats (APTs), who are typically a state, or state-sponsored group, and seek to steal data, or gain unauthorized access, for various purposes. Manufacturing companies in the United States are particularly vulnerable, as they have innovative developments that attract APT groups and often lack adequate security controls [16, 45].

# Design of Feedback for a System to Support Distance Project-Based Learning

Kosuke Sasaki[1,2][0000−0002−1011−8884] and Tomoo Inoue[3][0000−0003−3600−214X]

[1] Graduate School of Library, Information and Media Studies, University of Tsukuba, Ibaraki, Japan
[2] Faculty of Global Management, Chuo University, Tokyo, Japan
ksasaki@slis.tsukuba.ac.jp
[3] Institute of Library, Information and Media Science, University of Tsukuba, Ibaraki, Japan
inoue@slis.tsukuba.ac.jp

**Abstract.** This study focuses on project-based learning in a distance environment (distance PBL) and investigates the requirements for appropriate feedback that teachers provide to learners and the support system for teachers' feedback. In distance PBL, it is difficult for teachers to grasp the progress of individual learners, making it difficult for them to provide appropriate feedback, which is necessary for learners to proceed smoothly with their learning activities. Although previous studies have examined who should be given feedback, there has been a lack of consideration regarding what kind of feedback should be given to learners. The purpose of this study is to obtain insights into the appropriate feedback that teachers should give to learners in distance PBL by surveying previous studies. Based on these findings, this study presents six feedback requirements. This paper also presents an overview of a system to satisfy feedback requirements and to support teachers' feedback in distance PBL.

**Keywords:** Feedback · Assessment · Project-Based Learning · Distance Learning · Self-Regulation · Activity Report

## 1 Introduction

Distance learning has become increasingly popular. Distance learning has several advantages, including the possibility of taking classes from home without having to attend in-person school [15] and the possibility of learning at the learner's own pace using on-demand videos of classes [1, 17]. This study focuses on project-based learning (PBL) in a distant environment (distance PBL), in which research activities are the primary learning activities in higher education. PBL is a learner-centered learning method that integrates knowledge acquisition with activities aimed at solving real-world problems.

In PBL, it is difficult for teachers to devote sufficient time to each learner, as they generally teach multiple learners. Furthermore, it has been reported that

# Orientation-Dependent Cord Length Distribution Functions of Bounded Convex Domains ⋆

N. G. Aharonyan and V. K. Ohanyan[0000−0001−7029−2385]

American University of Armenia and Yerevan State University
victo@aua.am, victoohanyan@ysu.am, narine78@ysu.am

**Abstract.** In the last century German mathematician W. Blaschke formulated the problem of investigation of bounded convex domains in the plane using probabilistic methods. In particular, the problem of recognition bounded convex domains $\mathbb{D}$ by chord length distribution function (or density function).

**Keywords:** Stochastic and Integral geometry; Chord length distribution.

## 1 Introduction

The present paper continues the investigations begun in [11]. Random lines generate chords of random length in convex domain $\mathbb{D}$. The corresponding distribution (or density) function is called the chord length distribution function that we denote by $F(y)$ (or chord length density function $f(y)$). The form of the length density function is related to certain features of the corresponding figures. Poles of this function are related to parallel pieces of the contour and the form of $f(y)$ for $y$ close to its maximum is essentially related to smaller details of the contour. The determination of the chord length distribution function has a long tradition of application to collections of bounded convex bodies forming structures in metal and crystallography. The series of formulae for chord length distribution functions may be of use in finding suitable models when empirical distribution functions are given (see [9]).

In the initial stage of investigation mathematicians tried to find explicit expressions of the chord length distribution (or density) functions for concrete domains $\mathbb{D}$ in the terms of elementary functions. Till recently explicit expressions for the chord length distribution functions have been known in the case when $\mathbb{D}$ is a disc, a regular triangle (see [5]) and a rectangle (see [6]). These results have been obtained using the definition of chord length distribution function for a domain $\mathbb{D}$.

⋆ The investigation of the second author is done with partial support by the Mathematical Studies Center at Yerevan State University.

# Assessing Glaucoma Online Tools

Nelson Baloian[1][0000-0003-1608-6454] and Wolfram Luther[0000-0002-1245-7628]

[1] Department of Computer Science, University of Chile, Santiago, Chile
[2] Department of Computer Science, University of Duisburg-Essen, Germany

wolfram.luther@uni-due.de, nbaloian@dcc.uchile.cl

**Abstract.** In a recent publication, we presented online tools for computing the familial risk of stroke, for the occurrence of pathogenic variants in the BRCA1 or BRCA2 genes with impact on early breast and ovarian cancer disease, and for low- and high-stage prostate cancer [2, 3]. The forms collect information about the individuals, their specific disease patterns, medical examination results, and the lifestyle of the proband and his/her relatives. Furthermore, changes to the examination methods and redefinition of risk classes, data and model quality, as well as cross-cutting issues such as uncertainty and usability have been addressed. In this paper, we present various approaches, continuous or simple score-based risk models for estimating the 5-year risk that an individual with ocular hypertension will develop Primary Open Angle Glaucoma (POAG), the leading global cause of irreversible blindness. Finally, a customizable Dempster-Shafer (DS) risk assessment model is derived.

**Keywords:** Primary Open Angle Glaucoma Risk Calculator, Quality Metrics, Dempster-Shafer Risk Model Assessment.

## 1 Introduction

At least 2.2 billion people have a near or distance vision impairment. In almost half of these cases, the visual impairment could have been prevented or be remedied by preventive measures: early detection, periodical monitoring and appropriate treatment are crucial to avoid progressive and irreversible vision loss (within the next 5 years). But depending on the medical care available in their countries, often far too few people with distance vision problems due to refractive errors or a cataract have access to appropriate treatment [16].

The main cause of vision impairment is presbyopia, and to a far lesser extent:

- refractive errors
- cataract
- diabetic retinopathy
- glaucoma (Prevalence up to 111.8 million people worldwide by 2040, [12])
- age-related macular degeneration.

Glaucoma represents a degenerative optic neuropathy characterized by the progressive degeneration of retinal ganglion cells and the retinal nerve fiber layer, which leads to corresponding visual field defects [17]. There are four principal types: Primary Open-Angle Glaucoma (POAG) accounts for at least 90% of all cases–the angle between the iris and cornea remains open and drainage does not work properly–, followed by Closed-Angle, Congenital, a rare form of glaucoma in infancy, and Secondary Glaucoma. In addition to increased intraocular pressure, advanced age, patient's origin and

# Color Image Enhancement with Quaternion Fourier Transform-Based Alpha-Rooting

Anna A. Vardazaryan[1] and Artyom M. Grigoryan[2][0000−0001−6683−0064]

[1] Yerevan State University, Yerevan, Armenia
anna.vardazaryan3@edu.ysu.am
[2] ECE Dept, The University of Texas at San Antonio, San Antonio, TX 78249, USA
amgrigoryan@utsa.edu
https://ceid.utsa.edu/agrigoryan/

**Abstract.** This work presents a recent new effective method in color image enhancement method using the traditional non-commutative quaternion arithmetic for color images. In this arithmetic, the RGB color image together with the gray image is presented as a 4-component quaternion image. The method of alpha-rooting with the 2-D quaternion discrete Fourier transform (QDFT) for processing color images is described. Unlike traditional methods that enhance each color component individually, which often leads to color artifacts, the proposed method effectively processes all colors as one unit. The quaternion-based approach preserves the natural relationship among image components. The results of color image enhancement by the proposed method and comparison with the conventional color enhancement methods like color histogram equalization and channel-by-channel enhancement using the 2-D discrete Fourier transform-based alpha-rooting are described. Drone-captured images were utilized to showcase the practical application and effectiveness of our image enhancement method. The findings demonstrate that quaternion-based approach is more effective at preserving the color relationships and features of the image, offering a significant advancement in the field of image enhancement.

**Keywords:** Image enhancement · Alpha-rooting · Quaternion · Fourier transform.

## 1 Introduction

In the realm of digital image processing, image enhancement stands out as a crucial technique aimed at augmenting the quality of images for subsequent analysis or improved visual perception. Unlike general image processing, which encompasses a wide array of operations from acquisition to analysis, image enhancement is specifically tailored to modify the appearance of an image in a manner that is more pleasing to the observer or more suitable for analysis. Various factors can degrade image quality, including limitations of the capture device, poor lighting conditions, night scenes, dark objects, bright backgrounds, and other environmental challenges. As a result, essential details within images often

# Novel Gradient-Based Retinex Method for Image Enhancement

Armine A. Bayramyan[1] and Artyom M. Grigoryan[2][0000−0001−6683−0064]

[1] Yerevan State University, Yerevan, Armenia
armine.bayramyan@edu.ysu.am
[2] ECE Dept, The University of Texas at San Antonio, San Antonio, TX 78249, USA
amgrigoryan@utsa.edu
https://ceid.utsa.edu/agrigoryan/

**Abstract.** The retinex method, which includes single and multi-scale algorithms, is an effective method for grayscale and color image enhancement that proved color constantly and dynamic range compression. There are different implementations of the retinex algorithm, which allow different degrees of user control over parameters, color correction, intermediate steps and filters, and different forms of application. This study introduces a novel enhancement technique known as gradient-based retinex (GB-Retinex), which is uniquely applied to overcome common imaging challenges such as inconsistent lighting and unclear details. The presented method utilizes the Retinex theory, applying it to the low-pass filtered images by means of gradient operators (e.g. symmetric Laplacian gradients) to highlight critical features and enhance contrast. Through comparative analysis with the traditional gradient-based histogram equalization technique, GB-Retinex has shown to provide superior image improvements (we also use the Enhancement Measure Estimation (EME) for the comparison of methods). The enhanced clarity and detail achieved with GB-Retinex make it a preferable choice for image enhancement across different contexts. This advancement not only betters the visual quality but also the functional utility of images for subsequent analysis tasks. Illustrative examples with different images are given together with enhancement by method GB-HE.

**Keywords:** Image enhancement · Histogram equalization · Retinex · Gradients.

## 1 Introduction

Image enhancement is a crucial technique in digital image processing that aims to better the visual quality of images for human perception and more efficient computer analysis [1],[2]. It involves a range of methods, such as adjusting brightness, color correction, noise reduction, and sharpening details. These techniques are essential across various sectors; drones use enhanced images for clearer aerial views in agriculture and disaster relief, while thermal imaging is vital in medicine, building inspections, and security for heat detection. Enhanced medical images like X-rays and MRIs facilitate better diagnosis and treatment [3].

# Fairness in the Use of Medical Online Tools

Wolfram Luther [0000-0002-1245-7628] and Ashot Harutyunyan[0000-0003-2707-1039]
University of Duisburg-Essen, Department of Computer Science, Germany
Yerevan State University, Machine Learning Lab, and
Institute for Informatics and Automation Problems NAS RA, Armenia
wolfram.luther@uni-due.de harutyunyan.ashot@ysu.am

**Abstract.** The concept of fairness in the development and use of medical risk assessment tools is presented in this paper. After considering various approaches to a general definition of algorithmic fairness from the perspective of the implied sciences, guidelines for system requirements are formulated to highlight the different forms of fairness and their biases. These are for example poor data quality, inadequate models, bad accuracy and performance of algorithms or insufficient interaction or collaboration of stakeholders. The requirements are illustrated using the example of numerous tools for estimating the 5-year risk that an individual with ocular hypertension will develop Primary Open Angle Glaucoma (POAG), the leading global cause of irreversible blindness.

**Keywords:** Artificial Intelligence, Bias, Algorithmic Fairness, Discrimination, Risk Prevention, Medical Tool, Explainable Artificial Intelligence.

## 1 Introduction

Artificial intelligence (AI) is a collective term for technologies that support and enhance human abilities in hearing and seeing, analyzing, deciding, communicating and acting. It is based on the comprehensive digitalization of systems and processes (cognitive systems and their digital twins) and requires complex computer-based system architectures and their interfaces. It makes sense to differentiate between partial, complete and extended AI systems and AI-based applications, depending on how extensively human properties of learning, thinking, reasoning and further communication with native languages, facial expressions and gestures between humans and AI systems, and cooperation are enabled and comprehensively supported. This requires large language models and currently huge distributed computing resources.

The tools and technologies used are diverse and depend on the areas in which the AI solutions are running. It is therefore not surprising that special requirements are placed on the results of the use of AI technologies. AI should be explainable, comprehensible and meet specified quality standards [13]. This requires international agreements on domain-based criteria and metrics, procedures for validating results and, if necessary, adapting AI systems through learning in cooperation with experts. Results, predictions, recommendations, and the general use of AI-based assistance systems in the medical sector in particular require interdisciplinary evaluation, assessment and cooperative decision-making when it comes to the specific treatment of patients, preventive or follow-up care. As is known from game theory or multi-objective optimization, it cannot be excluded that not all criteria can be met equally well and that there should be procedures for finding compromises or defining a balance [13].

Of course, all these modern development processes are characterized by economic interests, the systems use existing knowledge, the companies involved use marketable

# Exploring Design Aspects of an AI-supported Farming Platform

Arsen G. Mikayelyan[0009-0009-6936-2396] and Ashot N. Harutyunyan[0000-0003-2707-1039]

ML Lab, Yerevan State University, 0025 Yerevan, Armenia
arsen.mikayelyan@ysu.am, harutyunyan.ashot@ysu.am

**Abstract.** We are exploring design aspects and elements of a comprehensive platform for smart agriculture. Focusing on the main concepts and an earlier prototype functionality, we propose the related vision on building such a solution to modernize farming practices through the integration of cutting-edge artificial intelligence (AI) approaches. The platform will be consisting of a suite of innovative features aimed at optimizing crop cultivation, irrigation management, and strategic planning for agricultural enterprises and regulatory actors.

**Keywords:** Intelligent agriculture, ML analytics of agronomic data, Gen AI and LLMs, real-time recommendations for farming.

## 1 Concepts and System Design

As in all spheres of human activities and domains of business, modern agriculture as well increasingly relies on data-driven decision making, real-time analytics, and automated management. Prior works on using AI in the agriculture include various studies and technology solutions. However, the review paper by Spanaki et al. [1] indicates that the disruptive potential of AI in the agricultural sector in terms of research and operations are still in infancy. A recent paper [2] focuses on opportunities and challenges that AI-driven approaches imply for a sustainable development in African continent. Vendors like IntelinAir [3] apply image processing techniques to effectively monitor health of agronomic fields. There are also various data sets [4]-[8] that can be helpful for building ML models in this domain.

In our study, we think more of an approach which is “platform”-ic in nature, comprising key features like an AI-driven assistant for planting guidance, an intelligent irrigation scheduling system with real-time monitoring capabilities, and personalized diagnostics for crop cultivation. Through the utilization of state-of-the-art ML algorithms, as well as Gen AI capabilities and Large Language Models (LLMs), the platform might provide dynamic conversational interfaces and a question-answering system, data analysis, and advanced statistical modeling to empower users with actionable insights.

In a full implementation scenario, the hierarchical backend service (Fig. 1) ensures seamless scalability, accommodating the diverse needs of multi-group users while maintaining robust performance and reliability. The platform's user-friendly interface enables farmers to visualize water lines, plot dimensions, and crop-specific cultivation

# An Explainable Clustering Algorithm using Dempster-Shafer Theory

Ricardo Valdivia[1], Nelson Baloian[1][0000-0003-1608-6454], Maral Chahverdian[2], Aram Adamyan[2], Ashot N. Harutyunyan[3,4][0000-0003-2707-1039]

[1] Department of Computer Science, University of Chile, Santiago 8330111, Chile
[2] American University of Armenia
[3] ML Lab, Yerevan State University, 0025 Yerevan, Armenia
[4] Institute for Informatics and Automation Problems of NAS RA, 0014 Yerevan, Armenia

**Abstract.** Clustering is an unsupervised learning method aimed at identifying data sets with similar characteristics. The quality of a clustering model is often assessed by its validity rather than its accuracy, using measures such as the Rand Index and the Correlation Coefficient. Recently, there has been an increasing interest in creating not only valid but also interpretable clustering models. The proposed solution in this study involves a clustering algorithm that generates labels for data and using the Dempster-Shafer classifier, creates clear rules ensuring interpretability for users.

**Keywords:** Explainable ML/AI, Clustering, Classification, Dempster-Shafer Theory.

## 1 Introduction and Motivation

Interpretability [1,2] refers to the model's ability to enable a human user to understand the how and why behind the model's specific outcomes. Current clustering algorithms, like K-means, are favored for their simplicity and scalability, yet they are often viewed as "black box" due to their opaque results. This has led to a growing focus on understanding and interpreting clustering models, and in developing model explanation techniques, such as SHAP (SHapley Additive exPlanations) [8], to provide a clear understanding of how clustering results are produced. The proposed solution in this study involves the development of a clustering algorithm that generates labels for data and, using the DS (Dempster-Shafer) classifier [3], creates clear rules ensuring interpretability for users. The development occurs in two stages: selecting optimal labels for training and consolidating the clustering algorithm, including training and predicting with the DS classifier for each data point. The implemented DSClustering algorithm achieves an effective combination of clustering techniques with enhanced interpretation through the automatic generation of categorical rules and precise adjustments in the training process of the classifier. The algorithm stands out for its ability to provide reliable and comprehensible clustering results, enhancing transparency and trust in data-driven decision-making.

# Embedded Interpretable Regression using Dempster-Shafer Theory

Nelson Baloian[1][0000−0003−1608−6454], Edgar Davtyan[6][0009−0007−1356−2220], Karen Petrosyan[2][0009−0000−8038−7337], Arnak Poghosyan[3][0000−0002−6037−4851], Ashot Harutyunyan[4,5][0000−0003−2707−1039], and Sergio Penafiel[1][0000−0002−0025−7805]

[1] Department of Computer Science, University of Chile
{nbaloian,spenafie}@dcc.uchile.cl
[2] American University of Armenia karen_petrosyan2@edu.aua.am
[3] Institute of Mathematics NAS RA arnak@instmath.sci.am
[4] Institute for Informatics and Automation Problems NAS RA
[5] Yerevan State University harutyunyan.ashot@ysu.am
[6] Picsart edgar.davtyan@picsart.com

**Abstract.** This paper introduces an innovative approach to regression analysis by incorporating the Dempster-Shafer theory to enhance the interpretability and accuracy of regression models. Our method, Embedded Interpretable Regression (EI Regression), segments continuous output variables into discrete interpretable classes and applies targeted regression models to these classes. We demonstrate the potential and effectiveness of our approach through initial experimental validation, showing that our model achieves competitive accuracy compared to traditional regression methods while significantly improving the model's interpretability. This work contributes to interpretable machine learning and offers a practical framework for applying Dempster-Shafer's theory in predictive modeling.

**Keywords:** Embedded Interpretable Regression · Interpretable Machine Learning · Dempster-Shafer Theory · Uncertainty Modeling

## 1 Introduction

In the realm of machine learning, developing models that not only predict accurately, but also provide insights into their decision-making processes is critical. This necessity is particularly pronounced in fields where decisions have significant consequences, such as healthcare, financial forecasting, and legal assessments. Traditionally, there is a noticeable trade-off between the accuracy of a model and its interpretability. Highly accurate models, such as deep learning networks, often function as "black boxes," where the decision processes are obscure. Conversely, simpler models like decision trees or linear regressions offer clarity but usually at the cost of performance on complex tasks.

The growing demand for interpretable and accurate models has led to new research exploring methodologies that can bridge this gap. The objective is to

# Improving the DSGD Classifier with an Initialization Technique for Mass Assignment Functions

Aik G. Tarkhanyan[1][0009−0000−7015−111X] and Ashot N. Harutyunyan[2,3][0000−0003−2707−1039]

[1] Mathematics and Mechanics Department at Yerevan State University, 0025 Yerevan, Armenia
hayk.tarkhanyan@edu.ysu.am
[2] ML Lab at Yerevan State University, 0025 Yerevan, Armenia
[3] Institute for Informatics and Automation Problems NAS RA, 0014 Yerevan, Armenia
harutyunyan.ashot@ysu.am

**Abstract.** Several studies have shown that the Dempster–Shafer theory (DST) can be successfully applied to scenarios where model interpretability is essential. Although DST-based algorithms offer significant benefits, they do face challenges in terms of efficiency. We present a method for the Dempster-Shafer Gradient Descent (DSGD) algorithm that significantly reduces training time—by a factor of 2.1—and also reduces the uncertainty of each rule (a condition on features leading to a class label) by a factor of 1.4, while preserving accuracy comparable to other statistical classification techniques. Our main contribution is the introduction of a "confidence" level for each rule. Initially, we define the "representativeness" of a data point as the distance from its class's center. Afterward, each rule's *confidence* is calculated based on *representativeness* of data points it covers. This confidence is incorporated into the initialization of the corresponding Mass Assignment Function (MAF), providing a better starting point for the DSGD's optimizer and enabling faster, more effective convergence. The code is available at https://github.com/HaykTarkhanyan/DSGD-Enhanced.

**Keywords:** Dempster-Shafer Theory · Interpretability · Mass Assignment Functions · Classification.

## 1 Motivation

Dempster-Shafer theory [1] has emerged as a powerful framework for developing classification algorithms that prioritize interpretability. This theory provides a mathematical approach for combining evidence from different sources to calculate the probability of an event, utilizing Dempster's rule of combination. Peñafiel et al. [2] have demonstrated that algorithm combining Dempster-Shafer theory with optimization techniques can offer substantial explainability, even when employing a limited number of rules, without sacrificing accuracy. The algorithm is

# An Empirical Analysis of Feature Engineering for Dempster-Shafer Classifier as a Rule Validator

Aneta Baloyan[2], Alexander Aramyan[3], Nelson Baloian[1][0000−0003−1608−6454], Arnak Poghosyan[4][0000−0002−6037−4851], Ashot Harutyunyan[5,6][0000−0003−2707−1039], and Sergio Penafiel[1][0000−0002−0025−7805]

[1] Department of Computer Science, University of Chile
nbaloian,spenafie}@dcc.uchile.cl
[2] Metric, Armenia aneta.baloyan@gmail.com
[3] American University of Armenia alex.aramyan@proton.me
[4] Institute of Mathematics NAS RA arnak@instmath.sci.am
[5] Institute for Informatics and Automation Problems NAS RA
[6] Yerevan State University harutyunyan.ashot@ysu.am

**Abstract.** Explainable AI methods are increasingly attracting the practitioner's attention since they convey important information about the nature of the phenomena being studied. Rule-generating models are considered one of the most explainable ones, however, there are so far not many attempts aimed at evaluating the rules generated by this kind of models. This paper proposes using an explainable classifier based on the Dempster-Shafer (DS) plausibility theory as a rule-validating mechanism to assess the reliability of the rule sets generated by AI models. The DS theory enables combining evidence from various sources while dealing with conflicting or even contradictory information, identifying trustworthy rules with high belief correctness. Our empirical analysis evaluates the DS-based classifier's ability to learn possibly complex numeric feature interactions on synthetic datasets. Results show the model excels at numeric interactions like differences and ratios and performs well regardless of class imbalance, but struggles with imbalanced categorical data. We conclude by proposing future work including comparative result analysis against traditional methods and developing hybrid approaches for more robust but at the same time interpretable AI systems.

**Keywords:** Explainable AI · Dempster-Shafer theory · Rule extraction · Rule validation · Feature interactions · Interpretable machine learning.

## 1 Introduction

The rise of artificial intelligence (AI) has revolutionized various industries, enabling us to tackle complex problems with unprecedented efficiency and effectiveness. However, as AI models grow in complexity, their inner workings become increasingly opaque, making it challenging to understand and validate the reasoning behind their decisions. Explainable AI (XAI) [1] has emerged as a crucial field, aiming to shed light on these black-box models and enhance trust in their

# Interpretability of Machine Learning Models in the Insurance Sector

Anna Sargsyan[1]

[1]PLAT.AI, Yerevan, Armenia
sargsyan.anna.g@gmail.com

**Abstract.** Recently, machine learning models become increasingly complex, and their interpretability becomes a critical concern. This paper investigates the challenges and methodologies associated with enhancing the interpretability of ML models, with a specific focus on their application within the insurance domain. It considers the fact that there are different stakeholders that need to understand how the model operates, and their needs can be covered with different approaches.

**Keywords:** Machine Learning, Insurance Sector, Interpretability, Black-Box Algorithm, Model Interpretability, Real Life Insights, Risk Score

## 1 Introduction

The financial services industry is undergoing a rapid and disruptive digital and AI transformation, and the insurance sector is not an exclusion. In insurance, AI/ML has significant potential to help reduce protection gaps by improving the availability, affordability, and accessibility of insurance on the back of increased personalization and improved cost-efficiency. In insurance, AI is most commonly used in underwriting, claims processing, customer service and fraud detection [1]. Various studies indicate that complex models generally outperform Generalized Linear Models (GLMs) in terms of predictive accuracy [2]. However, while linear models provide a straightforward interpretation of their predictions, Black-Box models lack a universally applicable solution for interpretation. Moreover, different stakeholders within the insurance domain necessitate varying levels of interpretability, prompting ML Engineers to address all their distinct needs [3]. Consequently, this paper concentrates on delineating the requirements of each stakeholder and proposing the most suitable interpretability technique for each.

## 2 Research Methodology

In this research, practical application of available methodologies is emphasized using different perspectives of stakeholders in the insurance domain. Real-life data from the Armenian car insurance market is used. However, in order not to reveal any commercial secret, a subset of data and subset of significant variables was used for training models and interpreting results. The models trained are targeting the probability that the policyholder will experience a claim in the requested policy period, hence binary classification methods are applied.

Afterward, the predicted probabilities are turned into Risk Scores, which are used by actuaries to determine prices for each police. Both linear and Black-Box complex models