International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AutoML and Automated Machine Learning Pipelines

Dharani R[1], Bharathi P A[2], Anees Priyanka V[3], Maheshwari D[4], Ms. S. Abikayil Aarthi[5], Noora K M[6], Johara kani S.[7]

Kings College of Engineering, punalkulam, pudhukottai,622001,Tamil Nadu.
Sivarayar Garden 2nd Street,M.chavady,Thanjavur,613001,Tamil Nadu.

ABSTRACT:

Automated Machine Learning (AutoML) has emerged as a transformative technology that streamlines the process of developing machine learning models. This paper provides an overview of recent advancements in AutoML techniques and automated machine learning pipelines. We discuss the fundamentals of AutoML, components of automated machine learning pipelines, state-of-the-art platforms and frameworks, challenges and limitations, future directions, and practical applications. Furthermore, we navigate through the labyrinth of state-of-the-art AutoML platforms and frameworks, ranging from Auto-Sklearn to H2O AutoML and Google AutoML, unraveling their features, capabilities, and real-world applications. Through a meticulous examination of these platforms, we uncover the democratizing force of AutoML, which empowers both novice and seasoned practitioners to harness the power of machine learning without being encumbered by the complexities of traditional model development. Through this comprehensive review, we aim to provide insights into the current state of AutoML research and its implications for the future of machine learning.

Keywords:

1. AutoML
2. Automated Machine Learning
3. Machine Learning Pipelines
4. Hyperparameter Optimization
5. Model Selection
6. Feature Engineering
7. Auto-Sklearn
8. H2O AutoML
9. Google AutoML
10. Neural Architecture Search (NAS)
11. Domain-Specific AutoML
12. Scalability
13. Interpretability
14. Federated Learning
15. Ethical Considerations

1. Introduction

Automated Machine Learning (AutoML) represents a significant leap forward in the field of machine learning, offering a transformative approach to model development. Traditionally, building machine learning models involved a labor-intensive process that required domain expertise, computational resources, and a deep understanding of various algorithms and techniques. However, the advent of AutoML has revolutionized this paradigm by automating many aspects of the model development pipeline.

Background and Motivation for AutoML:

The rapid growth of data and the increasing demand for machine learning applications across industries have underscored the need for efficient and accessible model development tools. AutoML addresses this need by automating key tasks such as feature engineering, hyperparameter optimization, and model selection, thereby democratizing the process of building machine learning models. By reducing the barriers to entry and empowering users with limited expertise to leverage the power of machine learning, AutoML has garnered significant interest from both academia and industry.

Overview of Traditional Machine Learning Pipeline Development:

Traditional machine learning pipeline development typically involves a series of manual steps, including data preprocessing, feature engineering, model selection, hyperparameter tuning, and model evaluation. Each of these steps requires careful consideration and expertise to ensure the development of high-performing models. However, this manual approach is time-consuming, resource-intensive, and often requires domain-specific knowledge, limiting its accessibility to a broader audience.

Importance of Automated Machine Learning Pipelines:

Automated machine learning pipelines offer a more efficient and scalable alternative to traditional model development methods. By automating repetitive tasks and leveraging computational resources effectively, AutoML accelerates the model development process, enabling users to iterate quickly and explore a broader range of models. Moreover, AutoML systems can adapt to evolving datasets and requirements, making them well-suited for dynamic and rapidly changing environments.

Objectives and Scope of the Paper:

In this paper, we aim to provide a comprehensive overview of AutoML and its implications for the field of machine learning. We will delve into the fundamentals of AutoML, including its core components and methodologies. Furthermore, we will explore the state-of-the-art AutoML platforms and frameworks, highlighting their features, capabilities, and real-world applications. Additionally, we will discuss the challenges and limitations of AutoML and outline future directions and emerging trends in the field. Through this exploration, we seek to elucidate the transformative potential of AutoML and its role in shaping the future of machine learning.
Section headings should be left justified, bold, with the first letter capitalized and numbered consecutively, starting with the Introduction. Sub-section headings should be in capital and lower-case italic letters, numbered 1.1, 1.2, etc, and left justified, with second and subsequent lines indented. All headings should have a minimum of three text lines after them before a page or column break. Ensure the text area is not blank except for the last page.

2. AutoML Fundamentals

Introduction to AutoML:

Automated Machine Learning (AutoML) represents a paradigm shift in the field of machine learning, aiming to automate and streamline the process of developing machine learning models. At its core, AutoML seeks to democratize machine learning by making it accessible to a broader audience, including users with limited expertise in data science and machine learning. By automating key tasks such as data preprocessing, feature engineering, model selection, and hyperparameter optimization, AutoML empowers users to build high-quality machine learning models with minimal manual intervention.

Evolution of AutoML from Traditional Machine Learning Practices:

The evolution of AutoML can be traced back to the challenges inherent in traditional machine learning practices, which often require significant time, effort, and expertise to navigate. As the volume and complexity of data continue to grow, traditional approaches to model development have become increasingly impractical and unsustainable. AutoML emerged as a response to these challenges, leveraging advancements in computational power, algorithmic innovation, and optimization techniques to automate and optimize the model development process.

Key Concepts: Hyperparameter Optimization, Model Selection, Feature Engineering Automation:

Central to AutoML are several key concepts that underpin its functionality and effectiveness. Hyperparameter optimization involves the automatic tuning of model parameters to maximize performance metrics such as accuracy or predictive power. Model selection refers to the process of automatically choosing the best-performing model architecture and algorithm for a given dataset and task. Feature engineering automation involves automatically generating and selecting relevant features from raw data, reducing the need for manual feature engineering.

Challenges in Manual Machine Learning Pipeline Development:

Manual machine learning pipeline development poses several challenges that AutoML seeks to address. These challenges include:

1. Time and resource constraints: Manual model development requires significant time and computational resources, limiting scalability and efficiency.
2. Expertise requirements: Building high-quality machine learning models traditionally requires expertise in data preprocessing, feature engineering, and model selection, which may not be readily available.
3. Reproducibility and consistency: Manual model development processes are prone to human error and variability, leading to inconsistencies and reproducibility issues.

4. Exploration limitations: Manual model development often involves exploring a limited subset of possible models and configurations, potentially missing out on optimal solutions.

By automating these tasks and addressing these challenges, AutoML offers a compelling solution for accelerating and democratizing the model development process.

3. Components of Automated Machine Learning Pipelines

Automated machine learning pipelines consist of several interconnected components that work together to automate the end-to-end process of model development. In this section, we will explore the key components of automated machine learning pipelines and the techniques associated with each component.

Data Preprocessing and Feature Engineering Automation:

Data preprocessing plays a crucial role in preparing raw data for machine learning models. Automated machine learning pipelines leverage various techniques to automate data preprocessing and feature engineering tasks, including:

- Missing Value Imputation: Techniques such as mean imputation, median imputation, and iterative imputation are used to fill missing values in datasets automatically.
- Feature Scaling: Automated pipelines scale features to a uniform range to ensure that they contribute equally to model training. Techniques like min-max scaling and standardization are commonly employed for feature scaling.
- Encoding Categorical Variables: Categorical variables are encoded into numerical representations using techniques like one-hot encoding, label encoding, or target encoding.
- Automated Feature Selection and Extraction: Automated pipelines employ feature selection techniques such as recursive feature elimination (RFE), feature importance ranking, and dimensionality reduction algorithms like principal component analysis (PCA) to automatically select or extract relevant features from the dataset.

Hyperparameter Optimization Techniques:

Hyperparameter optimization aims to find the optimal set of hyperparameters for a machine learning model to maximize its performance on a given dataset. Automated machine learning pipelines utilize various techniques for hyperparameter optimization, including:

- Grid Search: Grid search systematically searches through a predefined grid of hyperparameters to find the combination that yields the best performance.
- Random Search: Random search samples hyperparameters randomly from predefined distributions, offering a more efficient alternative to grid search for high-dimensional hyperparameter spaces.
- Bayesian Optimization: Bayesian optimization employs probabilistic models to efficiently search for optimal hyperparameters by balancing exploration and exploitation.
- Evolutionary Algorithms and Genetic Algorithms: Evolutionary algorithms, inspired by natural selection, use evolutionary principles such as mutation, crossover, and selection to iteratively optimize hyperparameters.

Model Selection Strategies:

Model selection involves choosing the most suitable machine learning algorithm and architecture for a given dataset and task. Automated machine learning pipelines employ various strategies for model selection, including:

- Cross-Validation: Cross-validation techniques such as k-fold cross-validation and stratified cross-validation are used to evaluate model performance on multiple subsets of the data.
- Ensemble Methods: Ensemble methods combine predictions from multiple base models to improve overall performance. Techniques such as bagging, boosting, and stacking are commonly used in ensemble learning.
- Meta-Learning Approaches: Meta-learning approaches leverage metadata about the dataset and previous model performance to guide the selection of appropriate machine learning algorithms and architectures.
- By automating these components and leveraging advanced techniques, automated machine learning pipelines streamline the model development process and enable users to build high-quality machine learning models efficiently.

4. State-of-the-Art AutoML Platforms and Frameworks

Automated Machine Learning (AutoML) platforms have gained significant traction in recent years, offering powerful tools and frameworks to streamline the process of model development. In this section, we provide an overview of some of the most popular AutoML platforms, compare their features and capabilities, and explore their real-world applications across various domains.

Overview of Popular AutoML Platforms:

1. Auto-Sklearn: Auto-Sklearn is an open-source AutoML toolkit built on top of the popular machine learning library, Scikit-learn. It automates the process of algorithm selection, hyperparameter optimization, and model ensembling, allowing users to build high-quality machine learning models with minimal manual intervention.
2. H2O AutoML: H2O AutoML is part of the H2O.ai platform, offering a suite of machine learning algorithms and tools for data science and AI. H2O AutoML automates the process of model selection, hyperparameter tuning, and feature engineering, providing a user-friendly interface for building predictive models.
3. Google AutoML: Google AutoML is a cloud-based platform that offers a range of AutoML services for vision, natural language, translation, and structured data. It leverages Google's infrastructure and advanced machine learning algorithms to automate model development and deployment, making it accessible to a wide range of users.
4. TPOT (Tree-based Pipeline Optimization Tool): TPOT is an open-source AutoML library that uses genetic programming to optimize machine learning pipelines. It searches through a space of possible pipelines, including preprocessing steps, feature selection, and model selection, to find the best-performing pipeline for a given dataset.

Comparison of Features, Capabilities, and Performance Metrics:

Each AutoML platform offers unique features, capabilities, and performance metrics. Auto-Sklearn, for example, provides a comprehensive set of machine learning algorithms and hyperparameter optimization techniques, while Google AutoML offers pre-trained models and scalable cloud infrastructure. H2O AutoML emphasizes ease of use and scalability, making it suitable for both small-scale and enterprise-level applications. TPOT, on the other hand, focuses on optimizing machine learning pipelines using genetic programming, allowing users to explore a wide range of possible pipeline configurations.

Case Studies and Real-World Applications:

AutoML platforms have been deployed in various domains and industries, demonstrating their versatility and effectiveness in solving real-world problems. For example, in healthcare, AutoML platforms have been used to analyze medical images, predict patient outcomes, and optimize treatment plans. In finance, AutoML has been applied to fraud detection, risk assessment, and portfolio optimization. Similarly, in e-commerce, AutoML platforms have been used to personalize product recommendations, optimize pricing strategies, and improve customer satisfaction.

By leveraging the capabilities of AutoML platforms, organizations can accelerate their machine learning initiatives, reduce time-to-market, and unlock new opportunities for innovation and growth. However, it's essential to carefully evaluate the features and performance of different AutoML platforms to choose the one that best aligns with the specific requirements and objectives of the project.

5. Challenges and Limitations

Automated Machine Learning (AutoML) has revolutionized the landscape of machine learning model development, but it also presents several challenges and limitations that must be addressed to ensure its effectiveness and reliability. In this section, we delve into some of the key challenges and limitations associated with AutoML.

Scalability Issues with Large Datasets:

One of the primary challenges in AutoML is scalability, particularly when dealing with large datasets. As the volume of data increases, the computational resources and time required for training and evaluating models also escalate. AutoML algorithms must be designed to handle large datasets efficiently, optimizing resource utilization and minimizing computational overhead.

Interpretability and Explainability Challenges in Automated Pipelines:

Another significant challenge in AutoML is the lack of interpretability and explainability in automated pipelines. As models become increasingly complex and opaque, understanding how they arrive at predictions becomes more challenging. Interpretability and explainability are essential for building trust in machine learning models, especially in regulated industries such as healthcare and finance.

Computational Resource Requirements and Scalability:

AutoML algorithms often require substantial computational resources, including memory, processing power, and storage. Scaling AutoML pipelines to handle larger datasets and more complex models can strain available resources and lead to scalability issues. Efficient resource management and optimization techniques are crucial for ensuring the scalability of AutoML systems.

Overfitting and Model Selection Biases in Automated Model Building:

AutoML pipelines are susceptible to overfitting, where models perform well on training data but generalize poorly to unseen data. Overfitting can occur when the search space of possible models is not constrained effectively or when hyperparameter optimization techniques are not tuned appropriately. Moreover, model selection biases may arise.

Addressing these challenges and limitations requires a concerted effort from researchers, practitioners, and industry stakeholders By overcoming these challenges, AutoML can realize its full potential as a transformative technology that democratizes machine learning and accelerates innovation across diverse domains.

6. Future Directions and Emerging Trends

As Automated Machine Learning (AutoML) continues to evolve, several promising directions and emerging trends are shaping the future of the field. In this section, we explore some of these key areas and discuss their potential implications for the advancement of AutoML.

Advances in Neural Architecture Search (NAS):

Neural Architecture Search (NAS) has emerged as a powerful technique for automating the design of neural network architectures. Recent advancements in NAS have led to the development of more efficient search algorithms, such as reinforcement learning and evolutionary strategies, which can discover novel architectures with improved performance and efficiency. The integration of NAS techniques into AutoML frameworks promises to accelerate the development of state-of-the-art deep learning models across a wide range of applications.

Multi-Objective Optimization in AutoML:

Traditional AutoML approaches focus on optimizing a single objective, such as model accuracy or computational efficiency. However, real-world machine learning tasks often involve multiple conflicting objectives, such as accuracy, interpretability, and computational cost. Multi-objective optimization techniques aim to simultaneously optimize these competing objectives, enabling AutoML systems to find a balance between performance and other desirable properties. By incorporating multi-objective optimization into AutoML pipelines, researchers can develop more robust and versatile machine learning models.

Integration of Domain Knowledge into Automated Pipelines:

Domain knowledge plays a crucial role in machine learning model development, guiding feature selection, model selection, and evaluation. Integrating domain knowledge into automated pipelines can improve the performance and interpretability of AutoML models, especially in specialized domains with unique characteristics and constraints. Techniques such as knowledge distillation, domain-specific feature engineering, and expert-guided optimization can enhance the effectiveness of AutoML systems and enable them to tackle complex real-world problems more effectively.

Federated Learning and Privacy-Preserving AutoML Approaches:

With growing concerns about data privacy and security, federated learning has emerged as a promising approach for training machine learning models on decentralized data sources while preserving data privacy. Federated AutoML techniques extend this concept to automated model development, allowing organizations to collaborate and share knowledge without compromising sensitive data. Privacy-preserving AutoML approaches, such as secure multi-party computation and differential privacy, ensure that sensitive information remains protected throughout the model development process.

Ethical Considerations in AutoML Development and Deployment:

As AutoML technologies become more pervasive, it is essential to address ethical considerations related to fairness, transparency, accountability, and bias in model development and deployment. Ethical AutoML frameworks and guidelines can help ensure that automated pipelines prioritize ethical principles and mitigate potential harms, such as algorithmic bias and discrimination. By fostering responsible and ethical practices in AutoML research and implementation, stakeholders can build trust and promote the responsible use of AI technologies in society.

By embracing these future directions and emerging trends, the AutoML community can unlock new opportunities for innovation and address complex challenges in machine learning and artificial intelligence. Through collaboration and interdisciplinary research, researchers, practitioners, and policymakers can shape the future of AutoML and harness its full potential to drive positive societal impact.

7. Applications and Case Studies

Automated Machine Learning (AutoML) has found diverse applications across various domains, revolutionizing the way machine learning models are developed and deployed. In this section, we delve into the use cases of AutoML in different industries, present detailed case studies highlighting its effectiveness, and discuss evaluation metrics and performance benchmarks for AutoML applications.

Use Cases of AutoML in Various Domains:

1. Healthcare: AutoML is used for medical image analysis, disease diagnosis, patient outcome prediction, and personalized treatment recommendation systems.
2. Finance: AutoML applications include fraud detection, credit risk assessment, algorithmic trading, and customer segmentation for targeted marketing.
3. E-commerce: AutoML is employed for product recommendation systems, demand forecasting, customer churn prediction, and sentiment analysis of customer reviews.
4. Manufacturing: AutoML is utilized for predictive maintenance, quality control, supply chain optimization, and anomaly detection in manufacturing processes.
5. Telecommunications: AutoML applications include network optimization, customer churn prediction, call center analytics, and predictive maintenance of network infrastructure.
6. Transportation: AutoML is used for route optimization, demand forecasting, predictive maintenance of vehicles, and real-time traffic analysis.

Detailed Case Studies Demonstrating the Effectiveness of AutoML:

1. *Healthcare*: A case study demonstrates how AutoML is used to develop a predictive model for early detection of diabetic retinopathy from retinal images, achieving high accuracy and reducing the need for manual screening.
2. *Finance*: An AutoML pipeline is deployed to build a fraud detection system for credit card transactions, significantly reducing false positives and improving detection rates compared to traditional rule-based approaches.
3. *E-commerce*: AutoML is employed to develop a personalized product recommendation system based on user behavior and purchase history, resulting in increased sales and customer satisfaction.
4. *Manufacturing*: A predictive maintenance model is built using AutoML to identify equipment failures before they occur, minimizing downtime and optimizing maintenance schedules.
5. *Telecommunications*: AutoML is utilized to analyze network traffic patterns and predict network failures, enabling proactive maintenance and improving service reliability.

Evaluation Metrics and Performance Benchmarks for AutoML Applications:

Evaluation metrics for AutoML applications vary depending on the specific task and domain. Commonly used metrics include accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (AUC-ROC), mean squared error (MSE), and root mean squared error (RMSE). Performance benchmarks for AutoML applications are often established based on comparisons with baseline models or industry standards, such as benchmarks established by competitions like Kaggle or academic datasets like the UCI Machine Learning Repository.

By showcasing the diverse applications of AutoML and presenting detailed case studies, stakeholders can gain insights into its potential impact and effectiveness in solving real-world problems across different domains. Moreover, establishing evaluation metrics and performance benchmarks helps to assess the quality and reliability of AutoML solutions and guide future research and development efforts.

8. Implementation Considerations and Best Practices

Implementing Automated Machine Learning (AutoML) requires careful consideration of various factors to ensure successful integration into existing workflows and maximize the effectiveness of automated pipelines. In this section, we provide guidelines and best practices for selecting appropriate AutoML tools and frameworks, strategies for integration, and recommendations for model deployment, monitoring, and maintenance.

Guidelines for Selecting Appropriate AutoML Tools and Frameworks:

1. Assess Requirements and Objectives: Understand the specific requirements, objectives, and constraints of the machine learning project, including data characteristics, task complexity, and resource availability.
2. Evaluate Features and Capabilities: Evaluate the features, capabilities, and performance of different AutoML tools and frameworks, considering factors such as supported algorithms, scalability, ease of use, and customization options.
3. Consider Integration with Existing Infrastructure: Assess the compatibility and integration capabilities of AutoML tools with existing data infrastructure, libraries, and software systems.
4. Explore Community Support and Documentation: Consider the availability of community support, documentation, tutorials, and resources for the chosen AutoML tool or framework to facilitate learning and troubleshooting.

5. Trial and Evaluation: Conduct pilot projects or trials to evaluate the performance and suitability of AutoML solutions for the specific use case before full-scale implementation.

Strategies for Integrating Automated Machine Learning Pipelines into Existing Workflows:

1. Identify Integration Points: Identify key integration points within existing workflows where automated machine learning pipelines can be seamlessly integrated, such as data ingestion, preprocessing, and model deployment.
2. Establish Data Pipelines: Establish robust data pipelines to ensure consistent data quality, preprocessing, and feature engineering across the entire machine learning workflow.
3. Automate Model Training and Evaluation: Implement automated processes for model training, hyperparameter optimization, and evaluation to streamline the model development process and reduce manual intervention.
4. Collaboration and Knowledge Sharing: Foster collaboration and knowledge sharing among data scientists, domain experts, and stakeholders to leverage domain knowledge and insights throughout the machine learning lifecycle.
5. Iterative Improvement: Implement processes for iterative improvement and refinement of machine learning models based on feedback, new data, and changing requirements.

Recommendations for Model Deployment, Monitoring, and Maintenance:

1. Scalable Deployment Infrastructure: Deploy machine learning models on scalable and reliable infrastructure to support production-level workloads and ensure high availability and performance.
2. Continuous Monitoring: Implement monitoring and alerting systems to track model performance, data drift, and system health in real-time, enabling proactive detection and mitigation of issues.
3. Feedback Loop and Model Retraining: Establish a feedback loop to collect user feedback, monitor model performance, and trigger retraining of models periodically or in response to significant changes in data or environment.
4. Version Control and Reproducibility: Maintain version control of models, code, and data to ensure reproducibility, traceability, and accountability throughout the machine learning lifecycle.
5. Security and Compliance: Implement security measures and compliance standards to protect sensitive data, ensure data privacy, and comply with regulatory requirements, such as GDPR or HIPAA.

By following these implementation considerations and best practices, organizations can effectively leverage AutoML to accelerate model development, improve decision-making, and drive innovation across various domains and industries.

9. Conclusion

In this paper, we have delved into the realm of Automated Machine Learning (AutoML), exploring its significance, components, challenges, applications, and best practices. As we conclude, let's summarize the key findings, discuss future prospects and research directions, and reflect on the significance of automated machine learning pipelines in advancing machine learning research and applications.

Summary of Key Findings and Contributions:

Throughout this paper, we have highlighted the transformative impact of AutoML in democratizing machine learning capabilities and streamlining the model development process. We discussed the evolution of AutoML, its fundamental components such as data preprocessing, hyperparameter optimization, and model selection, as well as challenges such as scalability issues and interpretability concerns. We presented use cases and case studies across various domains, demonstrating the effectiveness of AutoML in solving real-world problems. Additionally, we provided implementation considerations and best practices for selecting, integrating, and deploying AutoML solutions.

Future Prospects and Research Directions in AutoML:

Looking ahead, the future of AutoML holds exciting prospects and research directions. Advancements in areas such as neural architecture search (NAS), multi-objective optimization, federated learning, and ethical considerations will continue to drive innovation in the field. Additionally, integrating domain knowledge, enhancing interpretability, and addressing scalability challenges will be critical for the widespread adoption of AutoML in diverse applications and industries.

Closing Remarks on the Significance of Automated Machine Learning Pipelines:

Automated machine learning pipelines represent a paradigm shift in machine learning research and applications, offering a systematic approach to model development that empowers users with varying levels of expertise to build high-quality models efficiently.

By automating repetitive tasks and leveraging advanced optimization techniques, AutoML accelerates the pace of innovation, fosters collaboration, and unlocks new opportunities for organizations to extract actionable insights from their data. As we continue to explore the frontiers of AutoML, it is imperative to maintain a commitment to responsible and ethical AI practices, ensuring that automated machine learning technologies are deployed in a manner that maximizes benefits while mitigating risks.

In conclusion, automated machine learning pipelines hold immense promise for advancing machine learning research and applications, driving innovation, and shaping the future of artificial intelligence. By embracing AutoML and embracing interdisciplinary collaboration, we can unlock new possibilities, solve complex challenges, and create a world where intelligent automation empowers individuals and organizations to thrive in the digital age.

REFERENCES

1. Feurer, M., Klein, A., Eggensperger, K., Springenberg, J. T., Blum, M., & Hutter, F. (2015). Efficient and robust automated machine learning. In Advances in neural information processing systems (pp. 2962-2970).

2. Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). (2019). Automated Machine Learning: Methods, Systems, Challenges (Vol. 4). Springer.

3. Olson, R. S., Urbanowicz, R. J., Andrews, P. C., Lavender, N. A., Kidd, L. C., & Moore, J. H. (2016). Automating biomedical data science through tree-based pipeline optimization. In Applications of evolutionary computation (pp. 123-137). Springer.

4. Thornton, C., Hutter, F., Hoos, H. H., & Leyton-Brown, K. (2013). Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 847-855).

5. Kotthoff, L., Thornton, C., Hoos, H. H., Hutter, F., & Leyton-Brown, K. (2017). Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. Journal of Machine Learning Research, 18(1), 826-830.

6. Ras, M., van Gerven, M. A., & Haselager, P. (2018). Explainable neural networks based on additive index models. IEEE transactions on neural networks and learning systems, 29(10), 4793-4803.

7. Wang, H., & Yeung, D. Y. (2018). Towards Bayesian deep learning: A framework and some existing methods. IEEE Transactions on Knowledge and Data Engineering, 30(12), 2249-2261.

8. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 785-794).

9. Wu, S., Bao, F. S., Hu, J., & Wang, H. (2020). Towards privacy-preserving automated machine learning. IEEE Transactions on Knowledge and Data Engineering.

10. Gijsberts, C. M., Groenewegen, K. A., Hoefer, I. E., Eijkemans, M. J., Asselbergs, F. W., Anderson, T. J., ... & Lonneke, W. (2016). Race/ethnic differences in the associations of the Framingham risk factors with carotid IMT and cardiovascular events. PloS one, 11(8), e0161045.