Automated Machine Learning: A Survey of Model Selection, Hyperparameter Optimization, and Pipeline Integration

www.surveyx.cn

Abstract

Automated Machine Learning (AutoML) revolutionizes machine learning by automating processes such as model selection, hyperparameter optimization, and pipeline construction, thus reducing the expertise required to develop high-quality models. This survey explores the core components of AutoML, highlighting its transformative impact across domains like healthcare and business, where it enhances accessibility to advanced analytics. Despite its potential, AutoML faces challenges in scalability, interpretability, and domain-specific adaptations. The survey underscores the importance of user-centric designs and ethical considerations, emphasizing the need for transparent, fair, and sustainable AutoML systems. Future directions include enhancing computational efficiency, integrating domain-specific knowledge, and improving user interaction to ensure robust, adaptable solutions. By addressing these challenges, AutoML can democratize machine learning, making it more accessible and efficient across diverse sectors.

1 Introduction

1.1 Definition and Importance of AutoML

Automated Machine Learning (AutoML) revolutionizes machine learning by automating critical processes such as model selection, hyperparameter optimization, and pipeline construction. This automation significantly lowers the entry barriers traditionally associated with machine learning, which often demands specialized knowledge and involves cumbersome iterative workflows [1]. By streamlining these processes, AutoML enhances accessibility and efficiency, particularly for non-expert users [2].

AutoML democratizes access to machine learning technologies, enabling individuals and organizations with limited data science expertise to leverage advanced predictive models. This is achieved through the integration of techniques from various machine learning sub-fields, allowing for model construction that optimizes user-defined criteria, such as predictive performance [3]. Frameworks like AutoGluon-Tabular exemplify this advancement by incorporating best practices and improving predictive accuracy, thereby offering robust solutions for diverse applications [4].

AutoML systems also optimize resource sharing across multiple users, as evidenced by the development of multi-device, multi-tenant algorithms [5]. This capability is essential for enhancing access to efficient model selection and ensuring scalability in resource-constrained environments. By automating complex workflows, AutoML enables practitioners to concentrate on higher-level analytical tasks [6].

The robustness of AutoML is illustrated by its ability to maintain performance under challenging conditions, such as handling dirty data [7]. This robustness is crucial for practical deployment across various industries, where data quality often presents significant challenges. AutoML thus significantly

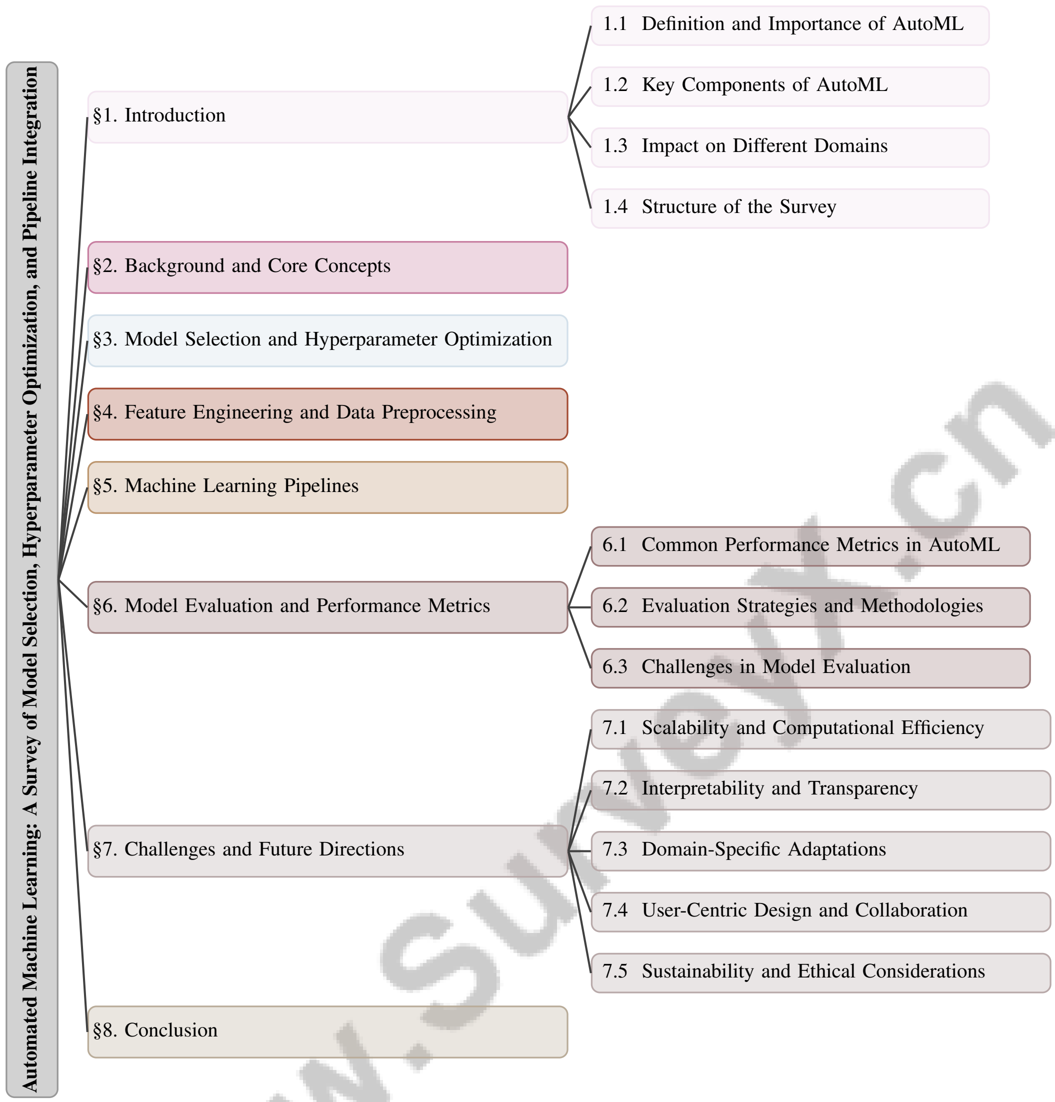


Figure 1: chapter structure

reduces the complexity and time needed to develop predictive models, making machine learning more accessible and efficient for a broader audience.

1.2 Key Components of AutoML

AutoML encompasses several core components that enhance the efficiency and accessibility of machine learning processes. Model selection is pivotal, automating the identification of the most suitable machine learning models for specific datasets, which is particularly beneficial for users lacking extensive expertise [8].

Hyperparameter optimization (HPO) focuses on fine-tuning model parameters to maximize performance. The integration of HPO with other AutoML tasks, such as neural architecture search, highlights the evolving capabilities of AutoML systems. For example, Auto-sklearn 2.0 employs portfolio-based model selection and budget allocation strategies to tackle challenges in existing frameworks [9].

Feature engineering, which includes data preprocessing and augmentation, is crucial for preparing data inputs that enhance model accuracy [4]. Despite advancements, current methods often necessitate user intervention, limiting accessibility for non-expert users [10].

Automating the entire machine learning lifecycle—from data preparation to model deployment—remains a significant challenge [6]. Frameworks like the AutoML-Agent exemplify efforts to automate these processes through a multi-agent system that collaborates efficiently across various stages [11].

Furthermore, AutoML systems must optimize multiple criteria without compromising model selection and preprocessing options [3]. The integration of domain knowledge poses additional challenges, as it constrains existing methods' ability to yield explainable and reliable outcomes.

The key components of AutoML—model selection, hyperparameter optimization, and feature engineering—are vital for its functionality, with ongoing research aimed at overcoming limitations and enhancing automation in machine learning processes. The inclusion of user collaboration through mixed-initiative approaches, as seen in systems like MILE, exemplifies the potential for tailored workflow optimization [1]. The GP-EI algorithm for resource allocation, which prioritizes model selection based on expected improvement rates and user needs, further reflects strategic advancements in AutoML [5].

1.3 Impact on Different Domains

AutoML is increasingly recognized for its transformative impact across various domains, significantly reducing the human effort required for machine learning tasks and enhancing accessibility to advanced analytics techniques [12]. In healthcare, AutoML advances applications such as diagnostics, treatment planning, patient management, and predictive analytics, which are crucial for improving patient outcomes and operational efficiencies [13]. Its integration in medical imaging further underscores its potential to drive advancements in this rapidly evolving field [14].

In business, AutoML is revolutionizing processes such as credit risk assessment and marketing strategies, enabling data-driven insights that support strategic decision-making [15]. The deployment of AutoML tools in these areas illustrates their capability to streamline complex analytical tasks, facilitating the adoption of AI-driven solutions in organizational functions. Similarly, in construction, AutoML addresses the complexity and inaccessibility of traditional machine learning methods, empowering professionals to utilize data effectively for informed decision-making [16].

Despite these advancements, implementing AutoML systems faces challenges related to transparency and user trust. Many frameworks operate as black boxes, leading to inefficiencies and potential user distrust due to a lack of transparency in model selection and optimization processes [17]. This issue is compounded by the risk of bias propagation within AI models, necessitating the development of interpretable feedback mechanisms that guide non-experts in improving model accuracy [18].

The adaptability of AutoML systems is further enhanced through the intersection with incremental learning, which improves models' ability to adapt to evolving data streams and concept drift [19]. This adaptability is essential for maintaining performance in dynamic environments and addressing real-world challenges. Additionally, AutoML's role in overcoming barriers across the machine learning lifecycle, particularly in deployment and monitoring, is significant for ensuring the sustained impact of machine learning technologies across various sectors [20].

Research has demonstrated AutoML's potential to streamline the training and optimization of large language models (LLMs), enhancing their efficiency and performance [21]. However, quantifying the environmental footprint of AutoML tools reveals significant challenges posed by high resource consumption and carbon emissions [22]. Addressing these environmental concerns is essential for sustaining the positive impact of AutoML across industries.

AutoML's influence across diverse fields is profound, offering substantial benefits in accessibility, efficiency, and the democratization of machine learning technologies. However, addressing challenges related to transparency, fairness, and user trust is crucial for maximizing AutoML's potential and ensuring its widespread adoption across industries [23].

1.4 Structure of the Survey

This survey provides an in-depth analysis of Automated Machine Learning (AutoML) and its components, emphasizing its critical role in enhancing accessibility for non-experts, improving the efficiency of machine learning workflows, and addressing challenges associated with human involvement in key stages. By introducing a classification system categorizing AutoML systems based on their level of automation, this review clarifies which aspects of the machine learning pipeline have been successfully automated and which still require manual intervention, outlining future research needed to advance end-to-end automation in machine learning processes [24, 25]. The paper begins with an **Introduction** defining AutoML and underscoring its importance, followed by an outline of key components such as model selection and hyperparameter optimization. The introduction also discusses AutoML's impact across various domains, setting the stage for subsequent exploration.

In **Section 2**, titled *Background and Core Concepts*, we delve into fundamental concepts underpinning AutoML, providing an overview of essential terms like model selection, hyperparameter optimization, feature engineering, data preprocessing, and machine learning pipelines. This section lays the groundwork for understanding the automation of these elements within AutoML frameworks.

Section 3, *Model Selection and Hyperparameter Optimization*, explores processes and strategies employed to automate model selection and hyperparameter tuning. This discussion emphasizes frameworks and tools that streamline AutoML, including innovative methodologies such as meta-learning for text representation and subset-based strategies optimizing computational efficiency, while highlighting emerging approaches that enhance text classification tasks through knowledge bases and large language models [26, 27, 28, 29, 30].

Section 4, *Feature Engineering and Data Preprocessing*, examines the roles of these tasks within AutoML, discussing automation challenges and methodologies developed to address them. This section emphasizes effective data preparation's importance in enhancing model performance.

focuses on , emphasizing their critical role in automating machine learning processes within AutoML systems. It explores how these pipelines integrate various tasks, streamline data preprocessing, model selection, and evaluation, ultimately enhancing the efficiency of deploying machine learning solutions. The discussion also highlights challenges associated with applying AutoML to unstructured data, such as text, necessitating robust pipeline designs accommodating diverse data types and user needs [31, 29, 25]. It covers the automatic assembly and optimization of pipeline structures, crucial for efficient workflows.

In , titled *Model Evaluation and Performance Metrics*, we provide a comprehensive overview of methodologies for assessing the performance of models produced by AutoML systems. This section covers common evaluation metrics, model assessment strategies, and specific challenges during evaluation. We highlight the complexities inherent in evaluating AutoML systems, which often involve multiple sub-components and require innovative methodologies to accurately capture user interactions and model effectiveness. Furthermore, we discuss the implications of our findings for future AutoML design and optimization, emphasizing the need for data-driven approaches to select appropriate evaluation strategies tailored to different datasets and objectives [29, 32, 33].

Section 7, titled *Challenges and Future Directions*, addresses current challenges faced by AutoML systems, such as scalability, interpretability, and domain-specific adaptations. The discussion highlights potential future research avenues, emphasizing the importance of user-centric design and addressing ethical considerations surrounding deployment and use [34, 35, 27, 36].

Finally, **Section 8**, *Conclusion*, summarizes the key points discussed throughout the survey, reflecting on AutoML's potential to democratize machine learning. This section reiterates the importance of addressing identified challenges to enhance the impact and adoption of AutoML technologies across various industries.The following sections are organized as shown in Figure 1.

2 Background and Core Concepts

2.1 Fundamental Concepts of AutoML

Automated Machine Learning (AutoML) enhances machine learning processes by automating tasks such as model selection, hyperparameter tuning, and data preprocessing, thus reducing the need for specialized expertise and broadening access to advanced analytics across various sectors [6]. AutoML systems aim to create robust workflows by sequencing preprocessing and predictive components while optimizing hyperparameters for improved performance [9].

Automating data preprocessing is crucial as it significantly impacts model performance, often more than hyperparameter tuning [10]. This underscores the importance of automating data-centric tasks, like feature selection and transformation, to enhance model accuracy and reliability. AutoML synthesizes workflows from a set of primitives, enabling complex tasks such as classification and regression [9].

The integration of AutoML with neural operations presents optimization challenges due to the high computational demands and multi-stage training processes [37]. Efficient resource management is essential to maintain model performance, particularly in addressing concept drift, where adaptive mechanisms are required to handle changes in data distribution over time [7].

Benchmarking is critical for evaluating AutoML systems across diverse datasets, facilitating framework comparisons and identifying strengths and weaknesses, especially in specialized domains with class imbalance and variable prevalence [38, 39]. Task descriptors enhance the evaluation of changes in AutoML systems, ensuring their suitability for real-world applications [6].

AutoML's potential is further demonstrated through the automated generation of machine learning pipelines, which streamline tasks traditionally dependent on expert knowledge for model design and tuning [1]. Zero-shot AutoML extends this potential by employing meta-learning to optimize model selection based on dataset characteristics without prior evaluations [39].

The fundamental concepts of AutoML focus on automating key machine learning tasks, addressing scalability and adaptability challenges, and integrating domain-specific knowledge to tailor solutions for specific tasks. These elements are vital for enhancing the accessibility and applicability of machine learning technologies, thereby democratizing advanced analytics across various sectors [6].

3 Model Selection and Hyperparameter Optimization

In Automated Machine Learning (AutoML), the processes of model selection and hyperparameter optimization are crucial for enhancing the performance of machine learning systems. As data complexity and task diversity increase, sophisticated methodologies are required to address these challenges. This section explores various frameworks and tools that facilitate model selection and hyperparameter optimization, improving both the efficiency and accessibility of machine learning technologies. Table 1 offers a comparative overview of key AutoML frameworks, emphasizing their optimization focus, automation level, and user accessibility.

Figure 2 illustrates the hierarchical organization of frameworks, tools, and innovative strategies in model selection and hyperparameter optimization within AutoML. This figure highlights the integration of various frameworks and tools that streamline machine learning processes, as well as innovative approaches that enhance the efficiency of model selection and hyperparameter optimization. By visualizing these relationships, the figure provides a clearer understanding of how these components interact to improve overall performance in AutoML.

3.1 Frameworks and Tools Implementing Model Selection and Hyperparameter Optimization

The AutoML ecosystem is enriched by frameworks and tools that automate model selection and hyperparameter optimization, streamlining complex workflows for users with varying levels of expertise. DriveML exemplifies this by allowing intricate machine learning processes to be executed through simple function calls, enhancing accessibility and efficiency [10]. Auto-sklearn 2.0 optimizes predictive performance across diverse datasets through portfolio-based strategies, automating the selection of optimization policies [9]. AutoGluon-Tabular demonstrates robustness and accuracy in handling tabular data, benchmarking favorably against other state-of-the-art frameworks [4], while many frameworks employ ensemble learning techniques and genetic programming to boost model performance.

The STREAMLINE framework emphasizes simplicity and transparency, initially focusing on binary classification with hyperparameter optimization across various algorithms [40]. This simplicity is vital for real-world AutoML adoption. AUTORUL integrates data cleaning, feature engineering, and regression modeling into a flexible pipeline optimized through Bayesian methods, exemplifying the automation of the entire data science process [41]. The DHPO method efficiently automates model selection and hyperparameter optimization in deep multi-target prediction settings, showcasing AutoML frameworks' adaptability to specialized tasks [42]. CLAMS-OT highlights innovative model selection by automating clustering algorithm selection based on dataset similarity using optimal transport distances [39].

AutoML-Agent integrates multi-agent systems, including various agents like Agent Manager and Prompt Agent, to automate model selection and hyperparameter optimization [11]. Additionally, a human-centered AutoML framework categorizes the lifecycle into stages, sub-tasks, user roles, and automation levels, enhancing user collaboration and satisfaction [6]. These diverse frameworks reflect strategic advancements in AutoML, enhancing workflow automation while integrating domain-specific knowledge and advanced learning strategies to optimize model performance across various domains [8].

As illustrated in Figure 3, which categorizes frameworks and tools within the AutoML ecosystem, model selection and hyperparameter optimization are essential for enhancing machine learning efficiency and performance. The figure highlights key tools that contribute to automating and enhancing machine learning processes across various domains. The first example contrasts various image classification methods, showcasing the performance of techniques like AutoSklearn and several GramML variants. The second example analyzes how hyperparameters such as optimizer, learning rate, embedding dimension, weight decay, and batch size affect a model's relative rank, crucial for model fine-tuning. The third example underscores the importance of hardware platforms in multi-objective functions, emphasizing the need to consider hardware capabilities in machine learning model design and deployment. These examples collectively demonstrate the intricate interplay between model selection, hyperparameter tuning, and hardware considerations in achieving superior machine learning outcomes [43, 44, 45].

3.2 Innovative Approaches in Model Selection and Hyperparameter Optimization

Recent advancements in AutoML have introduced innovative strategies that significantly enhance model selection and hyperparameter optimization, addressing the limitations of traditional methodologies. Zero-shot model selection for clustering uses optimal transport-based dataset similarity to enable automated recommendations without labeled data, facilitating efficient model selection in scenarios with scarce labeled data [39]. XD-Operations represent another innovative strategy, allowing a broader range of neural operations while utilizing successful existing architectures, crucial for adapting to diverse tasks and optimizing neural architectures within AutoML frameworks [37].

The cost-sensitive GP-EI algorithm optimizes resource allocation among multiple users, enhancing model selection and tuning, contrasting with traditional methods and making AutoML systems more efficient and scalable in multi-user environments [5]. Frameworks increasingly cater to both expert and non-expert users, addressing scalability and flexibility challenges [2]. This dual focus democratizes access to AutoML technologies, facilitating broader adoption.

These advancements aim to improve the robustness and versatility of AutoML solutions across diverse sectors, enhancing overall efficiency and effectiveness, steering the field toward more comprehensive and user-friendly solutions [24, 27, 29, 32]. The ongoing evolution of these strategies addresses key challenges such as transparency, interoperability, and flexibility in complex scenarios, ultimately enhancing the usability and effectiveness of AutoML frameworks.

Feature	DriveML	Auto-sklearn 2.0	AutoGluon-Tabular
Optimization Focus	Simple Function Calls	Portfolio-based Strategies	Tabular Data
Automation Level	High	Full Automation	Robust Automation
User Accessibility	Enhanced Accessibility	Diverse Datasets	Benchmarking Accuracy

Table 1: This table presents a comparative analysis of three prominent AutoML frameworks: DriveML, Auto-sklearn 2.0, and AutoGluon-Tabular. It highlights their distinct features, focusing on optimization strategies, automation levels, and user accessibility, thereby providing insights into their respective capabilities in enhancing machine learning processes.

4 Feature Engineering and Data Preprocessing

Feature engineering and data preprocessing are pivotal in Automated Machine Learning (AutoML), forming the groundwork for successful machine learning algorithm deployment. This section

AI-generated, for reference only.

delves into the challenges and methodologies related to automating these essential tasks, crucial for enhancing AutoML systems' efficiency and effectiveness.

4.1 Challenges in Automation

Automating feature engineering and data preprocessing in AutoML systems presents numerous challenges due to their inherent complexity and variability. A primary issue is the computational intensity of executing AutoML processes, which can hinder efficiency and scalability, especially with large datasets [10]. Overfitting hyperparameters to specific datasets poses another challenge, limiting generalization across diverse forecasting scenarios [38].

Current AutoML systems often lack user-centered design, failing to meet the needs of various user roles, particularly those with limited machine learning expertise, leading to usability issues [2]. Additionally, zero-shot model selection approaches may struggle if no analogous datasets exist within a meta-dataset [39]. Advanced operations like XD-Operations can increase computational costs, affecting training and inference times [37].

Transforming unstructured data into structured formats remains time-consuming and often requires expert intervention to ensure accuracy and efficiency [10]. Addressing these challenges is vital for advancing feature engineering and data preprocessing automation, thereby enhancing AutoML systems' robustness and applicability across diverse domains.

4.2 Methodologies and Tools

The automation of feature engineering and data preprocessing in AutoML systems significantly boosts machine learning models' efficiency and accuracy, enabling domain experts to access optimal solutions without extensive machine learning expertise. Although automation has advanced, critical tasks like understanding domain-specific data attributes and defining prediction problems still require human input, necessitating further research to enhance AutoML systems' autonomy [24, 46].

STREAMLINE automates data preparation, feature selection, modeling, and evaluation, ensuring reproducibility and transparency [40]. Maro enhances AutoML reliability by automatically debugging ML pipelines and generating remediated pipelines [47]. AutoML Trace visualizes collaborative processes between humans and AutoML systems, aiding in decision-making [48].

DiffPrep formulates preprocessing pipeline searches as a bi-level optimization problem, using gradient descent for efficient pipeline selection [49]. The dswizard framework improves hyperparameter optimization by adaptively pruning unpromising search space regions [50]. OutRank optimizes feature selection by leveraging a cardinality-aware variant of mutual information [51].

PipelineProfiler provides visual analytics within Jupyter Notebooks, aiding in understanding AutoML pipelines [52]. The CAAFE methodology uses large language models to iteratively generate features based on contextual information, enhancing feature engineering [53]. Traditional preprocessing techniques are integrated into the CNDPA method, underscoring feature engineering's ongoing relevance [54]. The AMLP toolkit simplifies pipeline creation and evaluation through symbolic representation [55].

These methodologies and tools illustrate diverse strategies for automating feature engineering and data preprocessing in AutoML systems. They enhance machine learning accessibility and effectiveness by minimizing manual data preparation and model training efforts, optimizing predictive model performance across tasks like text classification. Evaluations indicate that certain AutoML tools can outperform human data scientists in specific scenarios, streamlining workflows and improving machine learning outcomes [31, 25].

In Figure 4, feature engineering and data preprocessing are shown as critical steps in machine learning, significantly influencing predictive model performance. A comparative analysis of TPOT, AutoGluon, and AutoKeras highlights their strengths in ease of use, integration, and documentation, offering insights into effective feature engineering. The experimental setup for evaluating AutoML tools involves identifying potential tools, filtering based on configurations, and ensuring alignment with project requirements, optimizing feature engineering and data preprocessing phases [56, 57].

5 Machine Learning Pipelines

5.1 Concept and Importance of Machine Learning Pipelines

Machine learning pipelines provide structured frameworks that automate tasks from raw data transformation to model evaluation, encompassing data preprocessing, feature engineering, model selection, hyperparameter optimization, and deployment. These pipelines enhance reproducibility, efficiency, and scalability in machine learning projects [58]. Within Automated Machine Learning (AutoML), pipelines are crucial for automating complex tasks, allowing practitioners to focus on strategic decisions. The eTOP framework exemplifies this by optimizing resource use through conditional execution and early pipeline termination [59].

User-centered pipeline design further emphasizes their role in AutoML, accommodating user preferences and criteria like fairness and interpretability, thus democratizing machine learning technologies through intuitive interfaces [36]. Enhancements in pipeline interpretability and robustness, such as Robusta's automated feature selection, foster trust and transparency in machine learning applications [60]. Machine learning pipelines are indispensable for operationalizing AutoML, integrating components like data preprocessing and optimization techniques, thus making machine learning more accessible. Despite automation advancements, human expertise remains vital in understanding domain-specific data and defining prediction problems, highlighting the need for a user-automation partnership to meet diverse goals [24, 61, 25].

5.2 Automated Pipeline Assembly and Integration

Automated pipeline assembly and integration in AutoML systems enhance machine learning efficiency by streamlining complex workflows. This process accelerates modeling and optimizes data flow, crucial for managing increasing data complexity and volume. AutoML tools should foster a collaborative user-automation partnership, ensuring simplicity, reproducibility, and reliability [25, 62]. High-level symbolic expressions, like those in the AMLP toolkit, simplify pipeline optimization, allowing users to focus on strategic decisions [55].

Automated assembly integrates components like data preprocessing, feature engineering, model selection, and hyperparameter optimization into cohesive workflows adaptable to varying data and user needs. Meta-learning approaches and large language models enhance pipeline flexibility, optimizing data flows and model configurations for specific tasks [63, 11, 62, 50, 25]. Resource allocation and execution efficiency are optimized through strategies like conditional execution and early termination, as seen in the eTOP framework, and SubStrat's use of genetic algorithms for data subset selection, reducing execution times significantly [26, 27, 59, 49, 64].

Automated pipeline assembly and integration democratize machine learning access, enhancing efficiency and accelerating research. Despite challenges in achieving full automation, human expertise is essential for defining prediction problems and understanding domain-specific data, underscoring the need for a collaborative approach in AutoML tools [24, 25, 65, 36]. These systems facilitate complex workflow execution, enhancing machine learning accessibility and usability for a broader audience.

5.3 High-Level Pipeline Structure Optimization

High-level pipeline structure optimization in AutoML is vital for enhancing machine learning workflow efficiency, involving the selection and optimization of components like preprocessing methods, models, and hyperparameters, while considering their interactions. Recent advancements, such as novel neural architectures and hyper-hyperparameter selection strategies, highlight the potential for superior performance and accessibility in diverse applications [61, 66, 25, 32].

Adaptive search strategies, like those in the dswizard framework, dynamically adjust search spaces based on intermediate results, enhancing hyperparameter optimization by focusing resources on promising configurations [50]. Bi-level optimization techniques, as seen in DiffPrep, allow for extensive search space exploration, ensuring high-quality pipeline selection tailored to specific datasets and models [49]. User-centered pipeline design further underscores the importance of optimization, ensuring pipelines meet user needs through transparency and interpretability [6].

High-level pipeline optimization is essential for advancing AutoML by ensuring efficient, adaptable workflows. By leveraging sophisticated optimization methodologies and user-centered design, AutoML systems deliver robust, scalable machine learning solutions across domains. These advancements enhance machine learning accessibility for non-experts and address end-to-end pipeline complexities, enabling domain experts to engage effectively with machine learning processes. Ongoing research aims to automate critical steps, like hyperparameter tuning and model selection, improving efficiency and expanding AutoML's practical applications [24, 17, 46, 32].

6 Model Evaluation and Performance Metrics

In Automated Machine Learning (AutoML), model evaluation is essential for assessing performance across various tasks and datasets. This section delves into the intricacies of model evaluation, focusing on the selection of appropriate performance metrics and methodologies.

6.1 Common Performance Metrics in AutoML

Evaluating AutoML models requires diverse performance metrics to capture the complexities of model effectiveness across different tasks. For classification tasks, accuracy is a primary metric, providing a direct measure of classification capability, complemented by metrics like AUC and log-loss, especially when class distribution and misclassification costs are significant [4]. Regression tasks use metrics such as R^2 and RMSE to evaluate prediction accuracy and reliability, with macro-RRMSE and micro-RMSE addressing specific regression challenges [42]. In multi-label classification and multi-task learning, macro-AUPR and micro-AUPR offer a comprehensive view of a model's discriminative power across complex tasks [42].

Bayesian Optimization in AutoML frameworks involves performance measurement through iterative evaluations, focusing on the Pareto front to identify optimal configurations based on multiple performance criteria [3]. Operational metrics, such as execution time and resource utilization, are integral to AutoML performance, with frameworks like Auto-sklearn 2.0 emphasizing balanced error rates across repetitions [9]. The evaluation of execution time alongside accuracy, as demonstrated in DriveML, highlights the need for metrics capturing both efficiency and effectiveness [10].

Selecting performance metrics is crucial for ensuring models meet diverse user needs and application requirements. As AutoML evolves, developing sophisticated and context-aware evaluation metrics will be vital for advancing the field. Such metrics will enhance the reliability of automated solutions and address new hyper-hyperparameters introduced by recent AutoML systems. Automating these choices through data-driven methods, as shown in the analysis of 437 datasets from OpenML, can improve AutoML processes' efficiency and better support domain experts [31, 27, 32, 29, 24].

6.2 Evaluation Strategies and Methodologies

Benchmark	Size	Domain	Task Format	Metric
CCAI-AutoML[67]	59,904	Catalyst Discovery	Forecasting	RMSE, MAE
AutoML-UX[56]	5,000	Banking	Classification	Accuracy
AutoML-PD[68]	99,000	Phishing Detection	Binary Classification	Accuracy, AUC
TPOT[69]	10,307	Neuroimaging	Regression	Mean Accuracy Error
AutoMM[70]	55	Multimodal Learning	Classification	F1-weighted, R2
CMA-ES[71]	71	Classification	Classification	ROC AUC, Balanced Accuracy
AutoRecSys[72]	1,000,000	Recommender Systems	Rating Prediction	RMSE, MAE
AutoML-DL-CM[73]	11,320	Condition Monitoring	Classification	Accuracy

Table 2: This table provides a detailed overview of representative benchmarks used in AutoML evaluation, highlighting the diversity in dataset size, domain, task format, and evaluation metrics. Each benchmark is referenced with its corresponding study, illustrating the varied applications of AutoML systems across domains such as catalyst discovery, banking, and recommender systems.

Evaluation strategies and methodologies in AutoML are vital for ensuring model effectiveness and reliability across various tasks and datasets. A comprehensive framework often involves dissecting AutoML systems into sub-components for detailed analysis. The Evaluation Methodology for AutoML Systems (EMAS) advocates for an in-depth examination of AutoML systems through various metrics to enhance user engagement and system improvement [33]. Cross-validation techniques, commonly employed to assess classification errors and compare pipeline configurations, provide robust measures of model performance [55]. This minimizes the impact of data split variability, ensuring a more accurate assessment of model capabilities.

Standardized metrics across multiple datasets are essential for benchmarking AutoML frameworks, offering a consistent basis for comparing model performance. Table 2 presents a comprehensive summary of key benchmarks utilized in the evaluation of AutoML systems, emphasizing the importance of standardized metrics and diverse applications in enhancing the reliability and effectiveness of these systems. The OpenML AutoML Benchmark, encompassing 39 classification-focused datasets, exemplifies this approach by providing a comprehensive evaluation of different AutoML systems [74]. The GAMA framework illustrates an experimental setup where models are trained with diverse configurations, logging performance across multiple datasets to refine AutoML strategies [75]. Moreover, challenges of transparency and interpretability in AutoML must be addressed, as these qualities often hinder adoption and trust [76].

The performance of generated pipelines in AutoML is assessed through experiments measuring accuracy and efficiency across tasks, as demonstrated by the KGpip framework [77]. Additionally, employing data chunk evaluation across different types of concept drift measures the effectiveness of adaptation strategies through accuracy metrics [78]. Establishing comprehensive evaluation strategies is crucial for advancing AutoML, enabling data-driven decision-making for hyperparameters and evaluation strategies, ultimately enhancing AutoML system performance across applications [24, 27, 29, 32].

6.3 Challenges in Model Evaluation

Evaluating models in AutoML systems presents numerous challenges affecting their reliability and effectiveness. Overfitting is a significant concern, occurring when models are overly tailored to training data, leading to poor generalization on unseen data. The approach by Evans et al. enhances computational efficiency and prevents overfitting, improving performance on new datasets [79]. The complexity of integrating various tools and frameworks within AutoML pipelines complicates evaluation, necessitating continuous retraining to maintain robustness and explainability, as highlighted by Symeonidis et al. [80].

The lack of interpretability in AutoML models is a persistent challenge, as users often struggle to comprehend and trust system outputs. Jidney et al. emphasize developing standardized evaluation metrics and enhancing model interpretability to address these concerns [14]. Tools like XAutoML improve user understanding and trust by providing visual analytics that clarify model decisions [81]. Additionally, substantial computational resources required for evaluating AutoML systems can hinder accessibility and scalability. Kilickaya et al. note the absence of comprehensive methods for applying AutoML to incremental learning, impeding model evaluation in scenarios requiring continual adaptation [19].

Evaluation methodologies can also be limited by the availability and accuracy of task descriptors. Lorraine et al. highlight that unreliable task descriptors can compromise the evaluation process, necessitating the development of more robust task characterization methods [82]. Finally, challenges in documenting and monitoring model performance are highlighted by Alamin et al., who stress the need for improved documentation practices and community support to facilitate effective evaluation and monitoring [20]. Addressing these challenges is crucial for advancing AutoML and ensuring the reliability and applicability of automated machine learning solutions across diverse domains.

7 Challenges and Future Directions

7.1 Scalability and Computational Efficiency

Scalability and computational efficiency are central challenges in the evolution of Automated Machine Learning (AutoML), especially as these systems are applied to increasingly complex environments. Efficient strategies for navigating extensive search spaces, including feature engineering, algorithm selection, and hyperparameter tuning, are vital for enhancing performance [83]. Techniques such as subset-based strategies, exemplified by SubStrat, demonstrate the potential to reduce computation time while maintaining model accuracy, thereby reinforcing AutoML scalability [26]. Future research should focus on refining automated selection processes and exploring additional hyperparameter optimization to improve multi-target prediction performance [42].

Balancing accuracy and latency is crucial for AutoML scalability, with benchmarks emphasizing efficient resource use [38]. This balance is essential for practical applications, especially under limited computational resources [4]. Enhancing AutoML flexibility to accommodate diverse user needs, particularly for non-technical users, is also important [6]. Robustness evaluations in safety-critical applications and addressing coordination challenges across multiple devices can further improve scalability and resource allocation [7, 5].

Future research should integrate advanced algorithms and user interface improvements to expand existing frameworks for broader accessibility [10]. Optimizing computational efficiency in XD-Operations by broadening search spaces and addressing privacy and fairness concerns presents a promising avenue for enhancing AutoML adaptability [37]. Moreover, exploring computational efficiency improvements and alternative dataset similarity methods can bolster AutoML systems' robustness and applicability across domains [39].

7.2 Interpretability and Transparency

Interpretability and transparency are crucial for fostering trust and adoption of Automated Machine Learning (AutoML) systems. The "black box" nature of many models obscures decision-making processes, complicating validation and undermining user confidence [17]. This opacity is a significant barrier, particularly in high-stakes domains like healthcare, where understanding decisions is vital [14]. Ethical and practical challenges arise from AI applications lacking transparency, necessitating careful implementation to align with societal values [13].

Current research often overlooks transparency, flexibility, and domain knowledge, which are essential for effective AutoML implementation [12]. Enhancing transparency is crucial for building trust and enabling a deeper understanding of model decisions, especially in complex tasks like text classification [31]. Future research should prioritize rigorous evaluation frameworks to benchmark new methods, providing structured approaches to assessing transparency and interpretability [84]. Integrating user feedback and focusing on human-centered designs can improve engagement and system interpretability, allowing more intuitive interactions with automation processes [14]. Addressing these challenges can enhance AutoML transparency and interpretability, ultimately increasing trustworthiness and facilitating integration into diverse applications.

7.3 Domain-Specific Adaptations

Domain-specific adaptations in AutoML are essential for optimizing performance across various fields. Tailoring frameworks to specific domains can address unique challenges and leverage domain knowledge, enhancing model performance. This necessity is underscored by diverse data types and specific requirements across industries, such as audio, images, and complex datasets [85]. Incorporating feature engineering into AutoML frameworks is vital for improving performance, emphasizing domain-specific adaptations [86].

Robustness is crucial for developing domain-specific tools, especially with diverse datasets and fault types [7]. Future research should explore additional datasets and prioritize robustness to ensure reliable performance across domains. Refining policy selectors for out-of-distribution datasets and investigating alternative model selection strategies can enhance adaptability [9]. The demand for domain-specific adaptations is evident, as they significantly improve performance, usability, and applicability. Future inquiries should conduct comparative analyses of feature engineering and hyperparameter optimization across a broader array of libraries, prioritizing adaptation to unique domain requirements [24, 29, 63, 32].

7.4 User-Centric Design and Collaboration

User-centric design and collaboration are vital for democratizing access to AutoML technologies and enhancing usability across diverse user groups. Prioritizing user-friendly interfaces and collaborative features can make AutoML systems more accessible to individuals with varying technical expertise, facilitating broader adoption and effective human-machine interaction [87]. Future research should

AI-generated, for reference only.

emphasize developing systems that integrate seamlessly with existing frameworks, such as KGpip, which highlights the importance of collaboration and usability in hyperparameter optimization [77].

Incorporating human-centered design principles is crucial for addressing real-world challenges practitioners face [20]. Focusing on assistive automation tools and innovative data engineering techniques can develop systems that better integrate human expertise and decision-making [88]. This approach improves usability and ensures adaptability to specific user needs and preferences. The future of AutoML lies in creating systems that are fully automated, user-friendly, and capable of incorporating diverse criteria for model evaluation while supporting human decision-making [36]. Collaboration between the AutoML community and human-computer interaction researchers can guide development by user needs, resulting in more transparent and effective solutions [87].

Focusing on user-centric design and collaboration is pivotal for enhancing AutoML applicability and impact across domains. By addressing usability challenges and integrating human expertise, AutoML systems can become more intuitive and effective, facilitating broader access to advanced machine learning technologies. This improvement is crucial as current tools still require significant human involvement in critical stages, such as understanding domain-specific data attributes and defining prediction problems, limiting full automation potential. Advancing AutoML could empower domain experts and non-specialists to leverage machine learning without deep technical expertise, democratizing access to these powerful technologies [24, 31].

7.5 Sustainability and Ethical Considerations

Sustainability and ethical considerations are critical for responsible AutoML deployment across domains. Ethical considerations, particularly regarding data quality and balanced datasets, are essential for fair and effective outcomes. Ensuring interpretability and user-friendliness fosters trust and supports informed decision-making [5]. Sustainability is linked to model training efficiency, impacting resource consumption and environmental footprint. Optimizing execution efficiency, as demonstrated by frameworks like AutoGluon-Tabular, contributes to sustainability by minimizing resource use [4]. Energy-efficient strategies can mitigate environmental impact, especially in real-time and resource-intensive applications [5].

Ethical deployment involves addressing biases and promoting fairness, particularly in sensitive areas like healthcare. Emphasizing explainable AI and data-sharing practices responds to ethical concerns, aiming to improve transparency and accountability. Frameworks like STREAMLINE provide transparent, end-to-end pipelines, ensuring rigorous data analysis and comparison across algorithms. User-centered design in AutoML can mitigate fairness-related issues and enhance AI output reliability, fostering collaboration between human expertise and automated technologies [89, 90, 40, 2, 91]. Developing generalizable frameworks that adapt to various domains is critical for broad applicability and ethical deployment.

Ensuring fair access to computational resources is a significant ethical consideration. Discussions emphasize equitable access to resources in AutoML services [5]. Future research should refine meta-knowledge generation methods to enhance configuration space reduction strategies. Expanding candidate pipelines and benchmarks to include additional data sources and improve base model diversity are promising directions for advancing AutoML. Addressing sustainability and ethical considerations involves ensuring transparency, optimizing resource use, and enhancing user interaction. By prioritizing user-friendly interfaces, robust evaluation frameworks, and effective integration strategies, AutoML systems can be implemented ethically and efficiently. This approach addresses diverse stakeholder needs, enhancing usability for domain experts and facilitating seamless AutoML integration across industries [24, 21, 27, 32].

8 Conclusion

Automated Machine Learning (AutoML) has fundamentally transformed the landscape of machine learning by simplifying complex tasks like model selection, hyperparameter tuning, and pipeline integration. This survey highlights AutoML's capacity to replicate or surpass human performance in numerous machine learning applications, making it a valuable tool for both beginners and experienced practitioners. In fields such as healthcare, AutoML facilitates sophisticated data analysis, empowering professionals without deep data science expertise to extract meaningful insights.

However, the journey towards fully scalable AutoML systems that meet the diverse requirements of various industries is ongoing. Enhancing platform capabilities is imperative to achieve the desired level of self-service and widespread adoption. Automating the entire forecasting pipeline is crucial for boosting efficiency and minimizing the need for expert involvement. The development of interactive, fairness-aware AutoML systems is also essential, necessitating rigorous evaluation frameworks and strategies to address the challenges posed by unstructured data.

The necessity for interpretability and transparency in AutoML systems cannot be overstated, as these factors are critical for their widespread application. The integration of AutoML with large language models presents promising avenues for advancement, calling for innovative approaches to tackle emerging challenges. Additionally, strategies like SubStrat have shown promise in reducing computational demands while maintaining model accuracy, suggesting potential progress in AutoML methodologies.

Sustainability is another key consideration, with efforts focused on optimizing execution efficiency to minimize environmental impact. Balancing exploration and exploitation within extensive search spaces has proven effective in enabling AutoML frameworks to recommend optimal solutions. The human-in-the-loop model remains vital, ensuring that practitioners retain control and oversight, particularly during crucial phases of the automation process.

Despite significant progress in automating machine learning workflows, several challenges and unresolved questions persist. Future research should prioritize enhancing data processing and feature engineering automation to improve model reliability and performance. By addressing these issues, AutoML systems can become more robust, versatile, and applicable across a wide range of domains, ultimately fulfilling their potential to democratize machine learning.

References

[1] Doris Jung-Lin Lee and Stephen Macke. A human-in-the-loop perspective on automl: Milestones and the road ahead. *IEEE Data Engineering Bulletin*, 2020.

[2] Radwa Elshawi, Mohamed Maher, and Sherif Sakr. Automated machine learning: State-of-the-art and open challenges, 2019.

[3] Florian Pfisterer, Stefan Coors, Janek Thomas, and Bernd Bischl. Multi-objective automatic machine learning with autoxgboostmc, 2021.

[4] Nick Erickson, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li, and Alexander Smola. Autogluon-tabular: Robust and accurate automl for structured data, 2020.

[5] Chen Yu, Bojan Karlas, Jie Zhong, Ce Zhang, and Ji Liu. Automl from service provider's perspective: Multi-device, multi-tenant model selection with gp-ei, 2018.

[6] Dakuo Wang, Q. Vera Liao, Yunfeng Zhang, Udayan Khurana, Horst Samulowitz, Soya Park, Michael Muller, and Lisa Amini. How much automation does a data scientist want?, 2021.

[7] Tuomas Halvari, Jukka K. Nurminen, and Tommi Mikkonen. Testing the robustness of automl systems, 2020.

[8] Bin Liu. A very brief and critical discussion on automl, 2018.

[9] Matthias Feurer, Katharina Eggensperger, Stefan Falkner, Marius Lindauer, and Frank Hutter. Auto-sklearn 2.0: Hands-free automl via meta-learning, 2022.

[10] Sayan Putatunda, Dayananda Ubrangala, Kiran Rama, and Ravi Kondapalli. Driveml: An r package for driverless machine learning, 2021.

[11] Patara Trirat, Wonyong Jeong, and Sung Ju Hwang. Automl-agent: A multi-agent llm framework for full-pipeline automl, 2024.

[12] Kelly Azevedo, Luigi Quaranta, Fabio Calefato, and Marcos Kalinowski. A multivocal literature review on the benefits and limitations of automated machine learning tools, 2024.

[13] Alexandru-Ionut Imbrea. Automated machine learning techniques for data streams, 2021.

[14] Tasmia Tahmida Jidney, Angona Biswas, MD Abdullah Al Nasim, Ismail Hossain, Md Jahangir Alam, Sajedul Talukder, Mofazzal Hossain, and Md Azim Ullah. Automl systems for medical imaging, 2023.

[15] Marc Schmitt. Automated machine learning: Ai-driven decision making in business analytics, 2022.

[16] Rui Zhao, Zhongze Yang, Dong Liang, and Fan Xue. Automated machine learning in the smart construction era:significance and accessibility for industrial classification and regression tasks, 2023.

[17] Yuan Sun, Qiurong Song, Xinning Gui, Fenglong Ma, and Ting Wang. Automl in the wild: Obstacles, workarounds, and expectations, 2024.

[18] Behnaz Arzani, Kevin Hsieh, and Haoxian Chen. Interpret-able feedback for automl systems, 2021.

[19] Mert Kilickaya and Joaquin Vanschoren. What can automl do for continual learning?, 2023.

[20] Md Abdullah Al Alamin and Gias Uddin. Challenges and barriers of using low code software for machine learning, 2022.

[21] Alexander Tornede, Difan Deng, Theresa Eimer, Joseph Giovanelli, Aditya Mohan, Tim Ruhkopf, Sarah Segel, Daphne Theodorakopoulos, Tanja Tornede, Henning Wachsmuth, and Marius Lindauer. Automl in the age of large language models: Current challenges, future opportunities and risks, 2024.

[22] Tanja Tornede, Alexander Tornede, Jonas Hanselle, Marcel Wever, Felix Mohr, and Eyke Hüllermeier. Towards green automated machine learning: Status quo and future directions, 2023.

[23] Joseph D. Romano, Trang T. Le, Weixuan Fu, and Jason H. Moore. Is deep learning necessary for simple classification tasks?, 2020.

[24] Shubhra Kanti Karmaker Santu, Md. Mahadi Hassan, Micah J. Smith, Lei Xu, ChengXiang Zhai, and Kalyan Veeramachaneni. Automl to date and beyond: Challenges and opportunities, 2021.

[25] Doris Xin, Eva Yiwei Wu, Doris Jung-Lin Lee, Niloufar Salehi, and Aditya Parameswaran. Whither automl? understanding the role of automation in machine learning workflows, 2021.

[26] Teddy Lazebnik, Amit Somech, and Abraham Itzhak Weinberg. Substrat: A subset-based strategy for faster automl, 2022.

[27] Alexander Scriven, David Jacob Kedziora, Katarzyna Musial, and Bogdan Gabrys. The technological emergence of automl: A survey of performant software and applications in the context of industry, 2022.

[28] Jorge Madrid, Hugo Jair Escalante, and Eduardo Morales. Meta-learning of textual representations, 2019.

[29] Sebastian Brändle, Marc Hanussek, Matthias Blohm, and Maximilien Kintz. Evaluation of representation models for text classification with automl tools, 2021.

[30] Boshko Koloski, Senja Pollak, Roberto Navigli, and Blaž Škrlj. Automl-guided fusion of entity and llm-based representations for document classification, 2024.

[31] Matthias Blohm, Marc Hanussek, and Maximilien Kintz. Leveraging automated machine learning for text classification: Evaluation of automl tools and comparison with human performance, 2020.

[32] Matthias Feurer and Frank Hutter. Towards further automation in automl. In *ICML AutoML workshop*, page 13, 2018.

[33] Sungsoo Ray Hong, Sonia Castelo, Vito D'Orazio, Christopher Benthune, Aecio Santos, Scott Langevin, David Jonker, Enrico Bertini, and Juliana Freire. Towards evaluating exploratory model building process with automl systems, 2020.

[34] Amr Gomaa, Michael Sargious, and Antonio Krüger. Adaptoml-ux: An adaptive user-centered gui-based automl toolkit for non-ai experts and hci researchers, 2024.

[35] Marius Lindauer, Florian Karl, Anne Klier, Julia Moosbauer, Alexander Tornede, Andreas Mueller, Frank Hutter, Matthias Feurer, and Bernd Bischl. Position: A call to action for a human-centered automl paradigm, 2024.

[36] Florian Pfisterer, Janek Thomas, and Bernd Bischl. Towards human centered automl, 2019.

[37] Nicholas Roberts, Mikhail Khodak, Tri Dao, Liam Li, Christopher Ré, and Ameet Talwalkar. Rethinking neural operations for diverse tasks, 2021.

[38] David Hoffmann. Impact of hpo on automl forecasting ensembles, 2023.

[39] Prabhant Singh, Pieter Gijsbers, Murat Onur Yildirim, Elif Ceren Gok, and Joaquin Vanschoren. Clams: A system for zero-shot model selection for clustering, 2024.

[40] Ryan J. Urbanowicz, Robert Zhang, Yuhan Cui, and Pranshu Suri. Streamline: A simple, transparent, end-to-end automated machine learning pipeline facilitating data analysis and algorithm comparison, 2022.

[41] Marc-André Zöller, Fabian Mauthe, Peter Zeiler, Marius Lindauer, and Marco F. Huber. Automated machine learning for remaining useful life predictions, 2025.

[42] Dimitrios Iliadis, Marcel Wever, Bernard De Baets, and Willem Waegeman. Hyperparameter optimization in deep multi-target prediction, 2022.

[43] Hernán Ceferino Vázquez, Jorge Sanchez, and Rafael Carrascosa. Integrating hyperparameter search into model-free automl with context-free grammars, 2024.

[44] Yan Wen, Chen Gao, Lingling Yi, Liwei Qiu, Yaqing Wang, and Yong Li. Efficient and joint hyperparameter and architecture search for collaborative filtering, 2023.

[45] Daphne Theodorakopoulos, Frederic Stahl, and Marius Lindauer. Hyperparameter importance analysis for multi-objective automl, 2025.

[46] Yi-Wei Chen, Qingquan Song, and Xia Hu. Techniques for automated machine learning, 2019.

[47] Julian Dolby, Jason Tsay, and Martin Hirzel. Automatically debugging automl pipelines using maro: Ml automated remediation oracle (extended version), 2022.

[48] Jennifer Rogers, , and Anamaria Crisan. Tracing and visualizing human-ml/ai collaborative processes through artifacts of data work, 2023.

[49] Peng Li, Zhiyi Chen, Xu Chu, and Kexin Rong. Diffprep: Differentiable data preprocessing pipeline search for learning over tabular data, 2023.

[50] Marc-André Zöller, Tien-Dung Nguyen, and Marco F. Huber. Incremental search space construction for machine learning pipeline synthesis, 2021.

[51] Blaž Škrlj and Blaž Mramor. Outrank: Speeding up automl-based model search for large sparse data sets with cardinality-aware feature ranking, 2023.

[52] Jorge Piazentin Ono, Sonia Castelo, Roque Lopez, Enrico Bertini, Juliana Freire, and Claudio Silva. Pipelineprofiler: A visual analytics tool for the exploration of automl pipelines, 2020.

[53] Noah Hollmann, Samuel Müller, and Frank Hutter. Large language models for automated data science: Introducing caafe for context-aware automated feature engineering, 2023.

[54] Anish Thite, Mohan Dodda, Pulak Agarwal, and Jason Zutty. Concurrent neural tree and data preprocessing automl for image classification, 2022.

[55] Paulito P. Palmes, Akihiro Kishimoto, Radu Marinescu, Parikshit Ram, and Elizabeth Daly. Designing machine learning pipeline toolkit for automl surrogate modeling optimization, 2021.

[56] Oleg Bezrukavnikov and Rhema Linder. A neophyte with automl: Evaluating the promises of automatic machine learning tools, 2021.

[57] Forough Majidi, Moses Openja, Foutse Khomh, and Heng Li. An empirical study on the usage of automated machine learning tools, 2022.

[58] Daqin Luo, Chengjian Feng, Yuxuan Nong, and Yiqing Shen. Autom3l: An automated multimodal machine learning framework with large language models, 2024.

[59] Haoxiang Zhang, Juliana Freire, and Yash Garg. etop: Early termination of pipelines for faster training of automl systems, 2023.

[60] Xiaoyang Wang, Bo Li, Yibo Zhang, Bhavya Kailkhura, and Klara Nahrstedt. Robusta: Robust automl for feature selection via reinforcement learning, 2021.

[61] Sebastian Pineda Arango and Josif Grabocka. Deep pipeline embeddings for automl, 2023.

[62] Jiang Wu, Hongbo Wang, Chunhe Ni, Chenwei Zhang, and Wenran Lu. Data pipeline training: Integrating automl to optimize the data flow of machine learning models, 2024.

[63] Felix Neutatz, Marius Lindauer, and Ziawasch Abedjan. Automl in heavily constrained applications. *The VLDB Journal*, 33(4):957–979, 2024.

[64] Jinglue Xu, Jialong Li, Zhen Liu, Nagar Anthel Venkatesh Suryanarayanan, Guoyuan Zhou, Jia Guo, Hitoshi Iba, and Kenji Tei. Large language models synergize with automated machine learning, 2024.

[65] Anh Truong, Austin Walters, Jeremy Goodsitt, Keegan Hines, C. Bayan Bruss, and Reza Farivar. Towards automated machine learning: Evaluation and comparison of automl approaches and tools, 2019.

[66] Felix Mohr and Marcel Wever. Naive automated machine learning – a late baseline for automl, 2021.

[67] Renbo Tu, Nicholas Roberts, Vishak Prasad, Sibasis Nayak, Paarth Jain, Frederic Sala, Ganesh Ramakrishnan, Ameet Talwalkar, Willie Neiswanger, and Colin White. Automl for climate change: A call to action, 2022.

[68] Rizka Purwanto, Arindam Pal, Alan Blair, and Sanjay Jha. Man versus machine: Automl and human experts' role in phishing detection, 2021.

[69] Jessica Dafflon, Walter H. L Pinaya, Federico Turkheimer, James H. Cole, Robert Leech, Mathew A. Harris, Simon R. Cox, Heather C. Whalley, Andrew M. McIntosh, and Peter J. Hellyer. Analysis of an automated machine learning approach in brain predictive modelling: A data-driven approach to predict brain age from cortical anatomical measures, 2019.

[70] Zhiqiang Tang, Haoyang Fang, Su Zhou, Taojiannan Yang, Zihan Zhong, Tony Hu, Katrin Kirchhoff, and George Karypis. Autogluon-multimodal (automm): Supercharging multimodal automl with foundation models, 2024.

[71] Lennart Purucker and Joeran Beel. Cma-es for post hoc ensembling in automl: A great success and salvageable failure, 2023.

[72] Tobias Vente and Joeran Beel. The potential of automl for recommender systems, 2024.

[73] Payman Goodarzi, Andreas Schütze, and Tizian Schneider. Comparing automl and deep learning methods for condition monitoring using realistic validation scenarios, 2023.

[74] Stefan Coors, Daniel Schalk, Bernd Bischl, and David Rügamer. Automatic componentwise boosting: An interpretable automl system, 2021.

[75] Pieter Gijsbers and Joaquin Vanschoren. Gama: a general automated machine learning assistant, 2021.

[76] Hugo Jair Escalante. Automated machine learning – a brief review at the end of the early years, 2020.

[77] Mossad Helali, Essam Mansour, Ibrahim Abdelaziz, Julian Dolby, and Kavitha Srinivas. A scalable automl approach based on graph neural networks, 2022.

[78] Bilge Celik and Joaquin Vanschoren. Adaptation strategies for automated machine learning on evolving data, 2022.

[79] Benjamin Patrick Evans, Bing Xue, and Mengjie Zhang. Improving generalisation of automl systems with dynamic fitness evaluations, 2020.

[80] G. Symeonidis, E. Nerantzis, A. Kazakis, and G. A. Papakostas. Mlops – definitions, tools and challenges, 2022.

[81] Marc-André Zöller, Waldemar Titov, Thomas Schlegel, and Marco F. Huber. Xautoml: A visual analytics tool for understanding and validating automated machine learning, 2023.

[82] Jonathan Lorraine, Nihesh Anderson, Chansoo Lee, Quentin De Laroussilhe, and Mehadi Hassen. Task selection for automl system evaluation, 2022.

[83] Yang Li, Yu Shen, Wentao Zhang, Ce Zhang, and Bin Cui. Efficient end-to-end automl via scalable search space decomposition, 2022.

[84] Iddo Drori, Yamuna Krishnamurthy, Raoni Lourenco, Remi Rampin, Kyunghyun Cho, Claudio Silva, and Juliana Freire. Automatic machine learning by pipeline synthesis using model-based reinforcement learning and a grammar, 2019.

[85] Pierrick Pochelu, Serge G. Petiton, and Bruno Conche. A deep neural networks ensemble workflow from hyperparameter search to inference leveraging gpu clusters, 2022.

[86] Hassan Eldeeb, Mohamed Maher, Radwa Elshawi, and Sherif Sakr. Automlbench: A comprehensive experimental evaluation of automated machine learning frameworks, 2023.

[87] Thiloshon Nagarajah and Guhanathan Poravi. A review on automated machine learning (automl) systems. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pages 1–6. IEEE, 2019.

[88] Tijl De Bie, Luc De Raedt, José Hernández-Orallo, Holger H. Hoos, Padhraic Smyth, and Christopher K. I. Williams. Automating data science: Prospects and challenges, 2022.

[89] Jaimie Drozdal, Justin Weisz, Dakuo Wang, Gaurav Dass, Bingsheng Yao, Changruo Zhao, Michael Muller, Lin Ju, and Hui Su. Trust in automl: Exploring information needs for establishing trust in automated machine learning systems, 2020.

[90] Dakuo Wang, Justin D. Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. Human-ai collaboration in data science: Exploring data scientists' perceptions of automated ai, 2019.

[91] Hilde Weerts, Florian Pfisterer, Matthias Feurer, Katharina Eggensperger, Edward Bergman, Noor Awad, Joaquin Vanschoren, Mykola Pechenizkiy, Bernd Bischl, and Frank Hutter. Can fairness be automated? guidelines and opportunities for fairness-aware automl, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

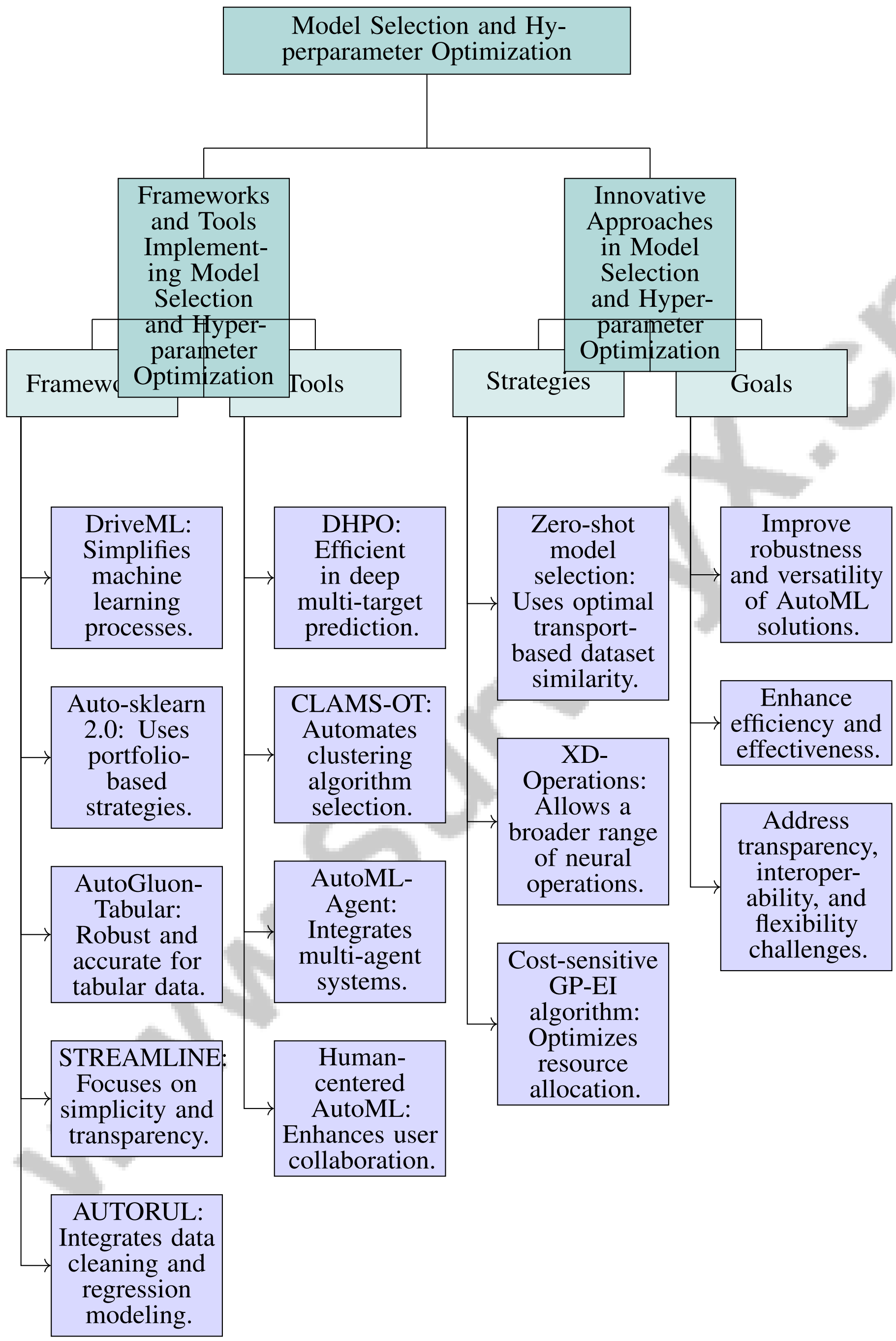


Figure 2: This figure illustrates the hierarchical organization of frameworks, tools, and innovative strategies in model selection and hyperparameter optimization within Automated Machine Learning (AutoML). It highlights the integration of various frameworks and tools that streamline machine learning processes and innovative approaches that enhance model selection and hyperparameter optimization efficiency.

AI-generated, for reference only.

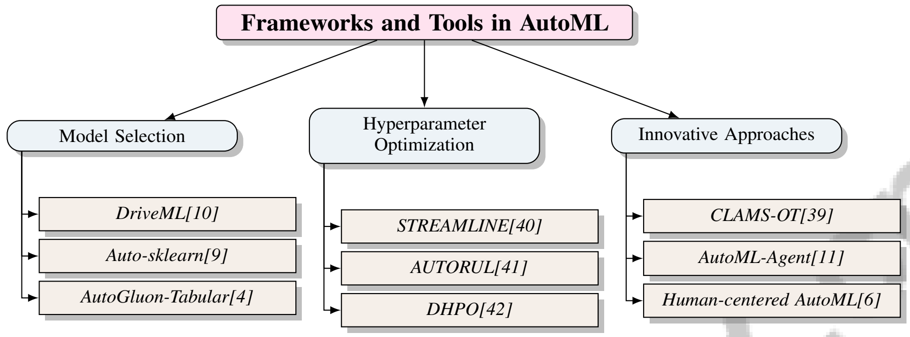
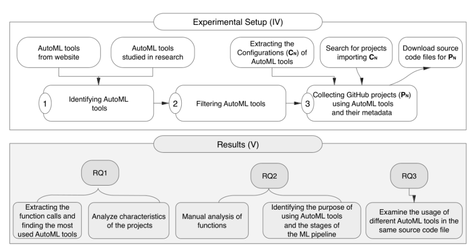


Figure 3: This figure illustrates the categorization of frameworks and tools within the AutoML ecosystem, focusing on model selection, hyperparameter optimization, and innovative approaches. Each category highlights key tools that contribute to automating and enhancing machine learning processes across various domains.

Tool	Simplicity of start	Integrations featured	Details in documentation
TPOT	Has a quick start tutorial for Classification and Regression on its GitHub page and in github.io documentation. Also features specific examples of usage.	Has a tutorial on work with Pytorch and Dask on github.io.	Has detailed documentation on Classification and Regression.
AutoGluon	Has a quick start tutorial for all main usecases on the official website. Has tabular quick start tutorial on the GitHub page.	Presents guidilines for pytorch users and explains work with H2O algorithms	Although it has documentation on some additional features, AutoGluon does not have detailed documentation on the main AutoML tools on its website. Has docs catalog on github, which might be used to generate documentation.
AutoKeras	Has a quick start tutorial for all main usecades on the official website. Has Image Classification quick start on the GitHub page.	Explains usage of TensorFlow Cloud with autokeras. Has a tutorial on Trains integration.	Has detailed documentation on all main classes and functions of the core features.

(a) Comparison of Different Machine Learning Tools: TPOT, AutoGluon, and AutoKeras[56]



(b) Experimental Setup for Analyzing AutoML Tools[57]

Figure 4: Examples of Methodologies and Tools