Forecasting in financial accounting with artificial intelligence – A systematic literature review and future research agenda

Marko Kureljusic
Chair of International Accounting, University of Duisburg-Essen - Campus Essen, Essen, Germany, and
Erik Karger
Chair of Information Systems and Strategic IT Management, University of Duisburg-Essen - Campus Essen, Essen, Germany

Abstract

Purpose – Accounting information systems are mainly rule-based, and data are usually available and well-structured. However, many accounting systems are yet to catch up with current technological developments. Thus, artificial intelligence (AI) in financial accounting is often applied only in pilot projects. Using AI-based forecasts in accounting enables proactive management and detailed analysis. However, thus far, there is little knowledge about which prediction models have already been evaluated for accounting problems. Given this lack of research, our study aims to summarize existing findings on how AI is used for forecasting purposes in financial accounting. Therefore, the authors aim to provide a comprehensive overview and agenda for future researchers to gain more generalizable knowledge.

Design/methodology/approach – The authors identify existing research on AI-based forecasting in financial accounting by conducting a systematic literature review. For this purpose, the authors used Scopus and Web of Science as scientific databases. The data collection resulted in a final sample size of 47 studies. These studies were analyzed regarding their forecasting purpose, sample size, period and applied machine learning algorithms.

Findings – The authors identified three application areas and presented details regarding the accuracy and AI methods used. Our findings show that sociotechnical and generalizable knowledge is still missing. Therefore, the authors also develop an open research agenda that future researchers can address to enable the more frequent and efficient use of AI-based forecasts in financial accounting.

Research limitations/implications – Owing to the rapid development of AI algorithms, our results can only provide an overview of the current state of research. Therefore, it is likely that new AI algorithms will be applied, which have not yet been covered in existing research. However, interested researchers can use our findings and future research agenda to develop this field further.

Practical implications – Given the high relevance of AI in financial accounting, our results have several implications and potential benefits for practitioners. First, the authors provide an overview of AI algorithms used in different accounting use cases. Based on this overview, companies can evaluate the AI algorithms that are most suitable for their practical needs. Second, practitioners can use our results as a benchmark of what prediction accuracy is achievable and should strive for. Finally, our study identified several blind spots in the research, such as ensuring employee acceptance of machine learning algorithms.

JEL Classification — C45, F37, M41

© Marko Kureljusic and Erik Karger. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/ licences/by/4.0/legalcode

The authors acknowledge the support by the Open Access Publication Fund of the University of Duisburg-Essen.
in companies. However, companies should consider this to implement AI in financial accounting successfully.

**Originality/value** – To the best of our knowledge, no study has yet been conducted that provided a comprehensive overview of AI-based forecasting in financial accounting. Given the high potential of AI in accounting, the authors aimed to bridge this research gap. Moreover, our cross-application view provides general insights into the superiority of specific algorithms.

**Keywords** Accounting, Forecasting, Artificial intelligence, Machine learning, Deep learning

**Paper type** Literature review

### 1. Introduction

Digital technologies have led to tremendous and fast-paced changes in almost all areas of life. Over the last three decades, advances in nanotechnology have enabled hardware development with steadily increasing computational power (Dingli *et al.*, 2021). As a result, digital developments like the Internet of Things or big data analytics are increasingly applied and used in many different areas. One of the most trending and hyped technologies of the digital age is artificial intelligence (AI) (Chollet, 2021). Driven by the aforementioned technological advances, AI has gained considerable attention and interest among managers and academics in recent years (Brock and von Wangenheim, 2019). Today, AI is used for many use cases, including speech or image recognition, medical diagnoses and automatizing routine labour (Goodfellow *et al.*, 2016). Given that, it is not surprising that AI is nowadays a multidisciplinary topic, which has implications for many different disciplines and industries.

Even though accounting is a traditional field with a long history, it has been subject to rapid changes in the past years that come along with the digital age (Berikol and Killi, 2021). In contrast to finance, which focuses on capital management, accounting uses transactional and imputed values to provide a true and fair view of companies’ assets, finances and income. While management accounting provides information to internal stakeholders, financial accounting provides information to external stakeholders (Penman, 2013). However, as harmonization efforts between the two subcategories have been observed in recent years, financial and management accounting are increasingly converging.

By changing workplaces and workflows, digital technologies lead to new opportunities in the accounting profession, but also demand new skills from the employees (see, for example, Leitner-Hanetseder *et al.*, 2021; Kruskopf *et al.*, 2020; Guthrie and Parker, 2016). Although accounting has always been subject to changes, current developments like the digital transformation and regulatory forces lead to developments that are more rapid and drastic than before (Hajkowicz *et al.*, 2016; Leitner-Hanetseder *et al.*, 2021). Apart from robotic process automation that already can help with routine tasks in accounting (Cooper *et al.*, 2019; Leitner-Hanetseder *et al.*, 2021), technologies like blockchain (Bonyuet, 2020; Maffei *et al.*, 2021), cloud computing (Huttunen *et al.*, 2019) and big data (Vasarhelyi *et al.*, 2015; Warren *et al.*, 2015) have a major influence on the future of accounting. Among all the digital technologies, AI is said to have a major influence on accounting, as it allows to identify patterns in large amounts of accounting data that can support firms’ decision-making as well as be used by stakeholders to conduct financial analyses (Lehner *et al.*, 2022). AI applied for accounting is a topic that has already been frequently discussed in research. Topics that are investigated are, among others, AI’s influence on the accounting profession (Stancu and Duțescu, 2021), its current limitations for accounting tasks (Losbichler and Lehner, 2021), or the influence of AI-based accounting on the accountant’s profession (Leitner-Hanetseder *et al.*, 2021).

Forecasting in accounting is an application area where AI-based algorithms are often and successfully used (Bertomeu, 2020; Kureljusic and Metz, 2023). The discipline of predicting future business events has a long tradition in accounting and has been investigated for decades. In the 1990s, the first ideas appeared that suggested the use of neural networks, a group of methods of AI, for accounting-related forecasting tasks. For example, neural networks
were used to predict quarterly accounting earnings (Callen et al., 1996), financial distress (Coats and Fant, 1991) or bankruptcy (Jo and Han, 1996). Since these pioneering works in the 1990s, AI and its use cases for accounting forecasting have seen tremendous research growth. New and more sophisticated AI methods emerged, and computing power, as well as available data volumes, increased. Likewise, the amount of research that uses AI for forecasting in accounting has grown steadily (Moll and Yigitbasioglu, 2019; Agostino et al., 2022).

AI-based forecasting for financial accounting tasks is a topic that is often and increasingly investigated. Despite the growing interest, the research within this field is dominated by the computer science or management discipline. These previous research projects mostly take a technical point of view and investigate specific AI algorithms and their accuracy for selected data sets, which often exclude outliers to achieve more accurate results. We believe that future researchers could make a wide variety of valuable contributions to research about AI-based forecasting within accounting. AI-related topics that so far are not investigated in accounting are, for example, user interactions with AI systems (Rzepka and Berger, 2018), the integration of AI with organizational strategies (Borges et al., 2021), or explainable AI (Meske et al., 2020).

Right now, to the best of our knowledge, there is no comprehensive review about AI-based forecasting in financial accounting that could serve as an initial understanding of this topic for future researchers. This lack of research is surprising, given the increasing relevance of AI for forecasting and the valuable contributions AI could provide. By conducting a systematic literature review, we aim to give a current overview of the research field of AI-based forecasting in financial accounting. This overview involves an investigation of the potentials, existing approaches and different use cases. From this goal, we derive our first research question:

**RQ1.** Which AI-based technologies and approaches are used for forecasting tasks in financial accounting, and what are the resulting benefits?

By integrating the findings and results from a broad range of various studies, a literature review helps to advance a research field (Snyder, 2019; Webster and Watson, 2002). We follow the argumentation of Paul and Criado (2020) and believe that a review article should “identify key research gaps” and serve as a platform for further research. We take a holistic perspective to outline what a future research agenda for underexplored topics might look like. Therefore, our second research question is as follows:

**RQ2.** How can future studies contribute to the field of AI-based forecasting in financial accounting, and what are promising future research questions and potential use cases?

The remainder of this article is structured as follows. In Section 2, we describe the foundations of financial accounting and AI. These foundations aim to equip the reader with the knowledge necessary to follow this article’s further results. After that, in Section 3, we explain the research method applied. Section 4 shows the results of our literature review. Section 5 will present a future research agenda derived from our findings in Section 4. Finally, Section 6 contains a discussion and concluding remarks.

### 2. Foundations

#### 2.1 Financial accounting

International accounting systems, such as International Financial Reporting Standards (IFRS) and United States Generally Accepted Accounting Principles (US GAAP), aim to present the actual net assets, financial position and results of a company’s operations. This information is included in the financial statements, primarily intended to provide potential and existing investors with information useful for decision-making (Penman, 2013). The characteristic
feature of accounting systems is their rule-based bookkeeping. Assets and liabilities are mostly recognized according to clearly defined criteria (Dai and Vasarhelyi, 2017). In this context, double-entry bookkeeping plays an elementary role in recording business transactions. Each business transaction is recorded both on the account and the offsetting account. The entry records are always balanced since each debit has a corresponding credit entry. The advantage of double-entry accounting over single-entry accounting is that transactions can be better tracked and verified (Sangster, 2016). This makes it easier for third parties, such as auditors and investors, to understand the business transactions.

However, double-entry accounting also has its limits. Ensuring the accounting equation of credits and debits does not guarantee that the correct accounts have been considered for accounting (Dai and Vasarhelyi, 2017). In addition, there is still a risk of fraudulent activities, as bookings can be modified or eliminated retrospectively (Faccia and Moşteanu, 2019). Recent fraud scandals such as Wirecard, Luckin Coffee or Steinhoff show that in some cases the accounting information presented does not correspond to the actual financial situation and can contain false statements (Raval, 2020). This leads to misallocations in the capital market, as investors make wrong investments based on false information (Giroux, 2008). Fraud in accounting is already a well-known problem, which has been tackled by numerous regulatory measures, such as the Sarbanes–Oxley Act, which arose due to the Enron scandal in 2001 (Coates, 2007). Regulatory adjustments are often retrospective and are made after a particular, usually negative, event. However, the current accounting scandals mentioned above show that fraud still exists and continues to be a major problem in accounting.

2.2 Artificial intelligence
Being coined for the first time in 1956, AI is one of the newest research fields within science and engineering (Russell et al., 2016). However, AI is not just computer science and mathematics but an interdisciplinary field with several significant contributions from other disciplines like economics, neuroscience and psychology (Taulli, 2019; Shalev-Shwartz and Ben-David, 2014). Searle (1980) was the first to differentiate between two different forms of AI: Strong and weak AI. Strong AI, also referred to as artificial general intelligence (Adams et al., 2012; van Gerven, 2017), can understand what is happening and might even be able to have emotions and feelings (Taulli, 2019). Strong AI aims to equip machines with human-like capabilities and intelligence (van Gerven, 2017). However, strong AI is not yet realized (Dingli et al., 2021).

Instead, today’s AI methods and applications are examples of weak AI. Weak AI systems are generally not intelligent and do not have emotions, feelings or a conscious mind (Searle, 1980; Taulli, 2019). Instead, weak AI applications are focused on single tasks. Therefore, they are developed only to behave instead of being intelligent (Franklin, 2014; Russell et al., 2016).

Nowadays, AI is a broad term covering many different technologies and approaches used for a wide variety of tasks. Machine learning (ML) is a group of AI methods among the most popular ones. The performance of ML improves with experience and aims to solve problems by using historical data or previous examples (Libbrecht and Noble, 2015). It is powering many aspects of modern society and is, for example, used to identify objects in images, which is increasingly present in consumer products like smartphones or cameras (LeCun et al., 2015). Depending on the way ML techniques learn, they can be broadly defined into two categories: Supervised and unsupervised learning techniques (Baştanlar and Ozuysal, 2014). Although other types of learning exist, for example, semi-supervised learning or online learning (Mohri et al., 2018), supervised and unsupervised learning are the most used and favoured ones (Alloghani et al., 2020).

The difference between supervised and unsupervised learning is the existence of labels in the training data (Alloghani et al., 2020). With supervised learning, the system receives labelled examples and input as the training data (Raschka and Mirjalili, 2019). In general
terms, ML can be defined as computational methods that use the experience to improve their performance or make more accurate predictions. The ML tool generates its experience from electronic data that is available to the system for analysis (Mohri et al., 2018). However, ML and AI algorithms are only as good as their training data (Dong and Rekatsinas, 2018; Halevy et al., 2009). Therefore, the quality and size of the data used to train the system are important factors (Goodfellow et al., 2016). Additionally, reinforcement learning is another learning paradigm that has received increasing attention in recent AI research. In comparison to supervised and unsupervised learning, reinforcement learning works a bit different. With reinforcement learning, an AI system (sometimes also referred to as agent) faces a problem and has to learn a certain behaviour to solve it. This behaviour is learned through trial-and-error interactions with the environment the agent is part of (Kaelbling et al., 1996). Hereby, the AI systems become their own teacher and do not require data, guidance or knowledge provided by humans (Silver et al., 2017).

Neural networks are one of the most frequently used AI technologies. Neural networks are inspired by the human brain’s structure and consist of small units that are connected with each other, called artificial neurons (Kureljusic and Reisch, 2022). These artificial neurons are small processing units that are connected with each other and generate an output based on learning rules and a received input. As such, neural networks aim to simulate brains in humans or other biological organisms (Aggarwal, 2018). In this context, deep learning is a term that is used to describe different types of complex neural networks. Due to the availability of large data sets and much computing power, deep learning grew significantly over the last years (Goodfellow et al., 2016). The architecture of deep learning comprises different modules or artificial neurons that are arranged in multiple layers. Each of these layers can transform input data and is able to learn. Deep learning has significantly improved many areas, including speech recognition, visual object recognition and object detection (LeCun et al., 2015).

2.3 Using artificial intelligence in financial accounting

Since accounting data is usually rule-based and well structured, they are well suited for automated evaluation using AI models. Especially financial key figures are useful for pattern recognition, as they are often related to each other (Soliman, 2008). In addition, due to the large number of variables reported in the balance sheet, the income statement and the cash flow statement, it is quite impossible to recognize all correlations without evaluations based on machines (Vlad and Vlad, 2021). Thus, it is conceivable that a negative long-term development (e.g. increased cost of materials) will be offset by a one-time positive effect (e.g. sale of property, plant and equipment). Accordingly, earnings alone are not a good indicator for forecasting a company’s future developments (Penman and Zhang, 2002). Nevertheless, an AI-based solution might be suitable for identifying complex relationships in accounting data and distinguishing short-term from long-term developments (Cho et al., 2020).

Forecasting is one of the use cases AI and ML techniques are frequently used for. Examples of forecasting include, among others, the forecasting of profit and loss variables (e.g. for revenues, see Kureljusic and Reisch (2022)) or cash flow variables (e.g. for customer payment dates, see Bahrami et al. (2020)). Different approaches can be used for forecasting tasks, namely classification, regression, ranking and clustering algorithms. With a classification approach, the problem consists of identifying categories for the investigated items (Baharudin et al., 2010). Ranking tasks aim at ordering items based on one or more criteria (Gerdes et al., 2021). Moreover, clustering involves dividing a set of elements into homogeneous subsets (Kansal et al., 2018). Finally, in contrast to the previous approaches, the regression outputs a continuous value that can be compared with other observations (Mohri et al., 2018).
The majority of the articles that are part of our literature review use AI and ML techniques that rely on supervised learning. There are only a few studies that apply algorithms that belong to unsupervised learning (Shi et al., 2009; Ding et al., 2019; Rainarli, 2019; Brown et al., 2020).

3. Research method

In this article, we aim to find an answer to the question of how AI-based algorithms are used for forecasting within accounting-related tasks. A literature review helps classify and integrate relevant findings from multiple disciplines by systematically collecting research. Therefore, we consider a systematic literature review as a useful method for generating a comprehensive overview of this research field that can serve as an initial overview and foundation for further studies. In conducting our systematic literature review, we follow the processes proposed by Kitchenham and Charters (2007) and Snyder (2019).

Although many different databases of scientific publications exist, we chose the Scopus and Web of Science (WoS), since these are among the largest ones (Forliano et al., 2021). The WoS contains scientific publications from 3,300 publishers and more than 12,000 journals (Mustak et al., 2021) and covers around 90 million documents (Forliano et al., 2021). Additionally, we used Scopus as a second database to cover a wider range of publications and minimize the risk of missing relevant literature. First, Scopus contains even more journals than the WoS (Paul and Criado, 2020). Furthermore, it has the advantage of not only covering journals or conference proceedings but also trade publications, books and different web sources (Kulkarni et al., 2009).

After the database selection, we constructed a search string. First, the search string needs to cover the technical part of AI-based technologies. Apart from artificial intelligence, we included machine learning and deep learning since these terms refer to two of the most popular AI-based methods. Furthermore, we included supervised, unsupervised and reinforcement learning as additional AI-related terms. The search string’s second part covers the use-case-related terms. In this study, we wanted to investigate how AI is used for forecasting in financial accounting. Therefore, we added forecasting, forecast, prediction and predicting to our search string. Since AI-based forecasting is used for many different tasks and goals, we considered it necessary to narrow our search down to accounting. Therefore, accounting was added at the end of the search string. This resulted in the following search string, which we applied in the databases mentioned above: (“artificial intelligence” OR “machine learning” OR “deep learning” OR “supervised learning” OR “unsupervised learning” OR “reinforcement learning”) AND (“forecast” OR “forecasting” OR “prediction” OR “predicting”) AND “accounting”

The search was conducted on 2 March 2022. We did not specify the time range and aimed to cover the eldest and newest publications. We searched the title, the abstract and the publication’s keywords. After eliminating duplicates, the initial sample consisted of 735 unique articles. Our first step was to exclude non-English articles. This led to the elimination of 12 publications. Since we wanted to focus only on peer-reviewed articles, our second step was to exclude non-peer-reviewed articles. This led to eliminating 58 further articles, with 665 publications remaining. Now, we started excluding articles based on the content. To be included for further investigation, the article had to deal with AI-based forecasts for financial accounting purposes. We started to investigate the article’s titles. Five hundred and six papers that did not fit our scope according to their title were eliminated, leaving 159 articles remaining. Of these articles, we then studied the keywords and abstracts. This led to the further elimination of 96 publications, remaining 63 publications. Our last exclusion step consisted of investigating the article’s full texts. Here, 16 further articles were considered to not fit our sample. As a final step, we completed our literature collection by applying...
backward searches to the remaining 47 articles. However, we could not find additional studies that were not already identified. This led to a final sample of 47 articles that were investigated. Figure 1 summarizes our applied literature collection steps.

4. Findings
The value of AI-based forecasting in financial accounting has already been identified by different studies (Ciampi et al., 2021; Bertomeu, 2020). After we received the final sample, our first step was to screen the content of the identified articles to get an idea about the content. Based on this initial assessment, we found that the existing research can be divided into three different categories. First, we will present the results dealing with bankruptcy predictions, as these forecasts are particularly important for the assessment of companies’ going concern assumptions. With 20 identified publications, this field has been addressed most often by prior research. In the finding’s second subsection, we deal with the field of financial analysis that is concerned with forecasting a company’s economic situation. Finally, the last subsection outlines fraud and error detection that are elementary for assessing the credibility of financial information.

4.1 Bankruptcy
The threat of insolvency or over-indebtedness is often seen as a major threat, especially by the company’s capital providers. In a corporate bankruptcy, they could lose their invested capital since the liquidator subordinates their claims. To avoid this misinvestment, equity and debt investors try to forecast a company’s future liquidity and financial situation (Agarwal and Taffler, 2008).

The earliest identified approach using AI to predict corporate bankruptcies has been proposed by Wilson and Sharda (1994). The authors compared the prediction accuracy of neural networks with that of multivariate discriminant analysis. Both prediction models are based on the supervised learning method, which allows for identifying patterns in the training data set, that can be used to forecast a company’s future liquidity. Their results show that neural networks have predicted significantly more accurately than multivariate discriminant analysis. Similar findings have been obtained by Lacher et al. (1995), who analyzed the same forecasting models for larger data sets with a longer time frame. In contrast to previous studies, Alici (1996) shows that Kohonen networks, as a method of

Source(s): Figure created by author
unsupervised learning, could also reliably distinguish solvent from insolvent companies. In another study by Kim (2005), the bankruptcy of companies was predicted by using more neurons per layer. However, this approach could not provide a higher prediction accuracy than previous studies.

A breakthrough was achieved by Tsai and Wu (2008), who were able to predict bankruptcy for Australian, German and Japanese companies more accurately. Unlike previous studies, more learning epochs and deeper neural networks were used. Besides the size of neural networks, Huang et al. (2008) were able to show that the accuracy of neural networks can be increased by calculating ratios from the input data. Another optimization option was proposed by Shi et al. (2009), showing that bagging as an ensemble learning method can also improve the prediction accuracy for predicting a company’s bankruptcy. Based on this, Lu et al. (2015) show that hybrid algorithms, such as using a support vector machine combined with particle swarm optimization, can also substantially improve the accuracy and robustness of bankruptcy predictions. The predictive power of support vector machines was also evaluated by Antunes et al. (2017) by comparing them with the logistic regression and Gaussian processes, with the result that Gaussian processes effectively improved the forecast performance. The fact that support vector machines could make more accurate bankruptcy forecasts than neural networks, decision trees, and logistic regression was also confirmed in the study by Rainarli (2019). The superior predictive power of support vector machines was also demonstrated in the study of Sehgal et al. (2021), who compared it with neural networks and logistic regression. Another algorithm for predicting a company’s bankruptcy is the random forest, which is generated from a large number of decision trees, and its predictive suitability has been tested by Kostopoulos et al. (2017), with the result that the Random Forest performed even better than the support vector machine.

Further developments of tree-based forecasting models are modern gradient boosting algorithms. One of these is the TreeNet algorithm, which adds another tree to correct the predicted error after each iteration. Jones and Wang (2019) have shown in their study that the TreeNet algorithm provides substantially more accurate bankruptcy forecasts than conventional models such as logistic regression. Besides ensemble and tree-based methods, the ant colony optimization algorithm is another way to implement swarm intelligence. Especially in optimization problems, ant colony optimization is useful to search for the simplest solution (e.g. reducing the rule complexity). For example, the study of Uthayakumar et al. (2020) found that the ant colony optimization algorithm provides better bankruptcy forecasts than logistic regression or the random forest, which had often performed very well in previous studies.

Further technological innovations, such as parallelization and distributed storage, have enabled neural networks to become larger and deeper. Thus, Alexandropoulos et al. (2019) found that deep neural networks can predict more accurately than the logistic regression and the naïve Bayes approach. Similar findings emerged from the study by Cao et al. (2020), observing the superior predictive ability of deep neural networks over support vector machines and Bayesian networks. Recent developments in deep learning are long-term short-term neural networks, which simulate a short-term memory by remembering previous expectations. In the study by Jang et al. (2020), it was shown that long-term short-term neural networks could provide very reliable bankruptcy forecasts if they receive macroeconomic and industry data in addition to financial company data.

For better classification of company data, a comparison with the peer group is often useful. Ding et al. (2019) present K-medians clustering, a method that can predict companies’ bankruptcy and misstatements based on a peer group cluster analysis. Besides K-medians clustering, the stacked autoencoder is another method of unsupervised learning. The study by Soui et al. (2020) proved that the stacked autoencoder could be used for feature extraction.
to increase the data quality of the input data. In addition, the data preprocessing significantly increased the models’ forecast accuracy.

In addition to comparative data from peer group companies and effective feature extraction, it is conceivable that textual disclosures can be used as additional input data for prediction models. Mai et al. (2019) have used textual data from companies’ annual reports together with traditional financial ratios and market-based variables. Their inclusion of text data leads to a significant improvement in the prediction accuracy of neural networks, which thus provide more accurate predictions than random forest or logistic regression. Table 1 presents an overview of all studies dealing with bankruptcy predictions regarding the evaluated models, sample sizes and accuracies. If multiple data sets were analyzed, the data set with the highest accuracy is shown.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Evaluated models</th>
<th>Best model</th>
<th>Sample size</th>
<th>Period</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilson/Sharda (1994)</td>
<td>MDA, NN</td>
<td>NN</td>
<td>129</td>
<td>1975–1982</td>
<td>97.50%</td>
</tr>
<tr>
<td>Lacher et al. (1995)</td>
<td>MDA, NN</td>
<td>Kohonen network</td>
<td>282</td>
<td>1970–1989</td>
<td>91.50%</td>
</tr>
<tr>
<td>Kim (2005)</td>
<td>DT, MDA, NN, Adapted NN (incl. ratio calculation of input features)</td>
<td>NN Adapted NN</td>
<td>654</td>
<td>N.A.</td>
<td>80.00%</td>
</tr>
<tr>
<td>Huang et al. (2008)</td>
<td>NN ensemble</td>
<td>NN ensemble</td>
<td>690</td>
<td>2004</td>
<td>97.87%</td>
</tr>
<tr>
<td>Tsai/Wu (2008)</td>
<td>DT, NN, NN, Adapted NN (incl. bagging), Nearest Neighbour, SVM, ZeroR</td>
<td>NN</td>
<td>1,000</td>
<td>N.A.</td>
<td>75.60%</td>
</tr>
<tr>
<td>Shi et al. (2009)</td>
<td>NN ensemble</td>
<td>NN ensemble</td>
<td>690</td>
<td>2004</td>
<td>97.32%</td>
</tr>
<tr>
<td>Lu et al. (2015)</td>
<td>SVM, Hybrid model (SVM + SPSO)</td>
<td>Hybrid model</td>
<td>250</td>
<td>N.A.</td>
<td>99.21%</td>
</tr>
<tr>
<td>Antunes et al. (2017)</td>
<td>Gaussian process, LR, SVM</td>
<td>RF</td>
<td>2,000</td>
<td>2002–2006</td>
<td>98.13%</td>
</tr>
<tr>
<td>Kostopoulos et al. (2017)</td>
<td>BN, NN, RF, SVM</td>
<td>SVM</td>
<td>435</td>
<td>2003–2005</td>
<td>70.19%</td>
</tr>
<tr>
<td>Alexandropoulos et al. (2019)</td>
<td>LR, NB, NN</td>
<td>NN</td>
<td>450</td>
<td>2003–2004</td>
<td>73.20%</td>
</tr>
<tr>
<td>Jones/Wang (2019)</td>
<td>LR, TreeNet</td>
<td>CNN, Adapted CNN (incl. word embedding), LR, RF, SVM</td>
<td>492,227</td>
<td>2014–2018</td>
<td>85.60%</td>
</tr>
<tr>
<td>Mai et al. (2019)</td>
<td>LR, NN, RF, SVM</td>
<td>SVM</td>
<td>120</td>
<td>2008–2014</td>
<td>85.83%</td>
</tr>
<tr>
<td>Rainarli (2019)</td>
<td>DT, LR, NB, Nearest Neighbour, NN, SVM, ZeroR</td>
<td>SVM</td>
<td>1,563,010</td>
<td>1961–2018</td>
<td>83.72%</td>
</tr>
<tr>
<td>Cao et al. (2020)</td>
<td>BN, DT, LR, NN, SVM</td>
<td>LSTM</td>
<td>1,378</td>
<td>1980–2016</td>
<td>98.54%</td>
</tr>
<tr>
<td>Jang et al. (2020)</td>
<td>AB, LD, LR, NN, RF, SAE, SVM, XGBoost</td>
<td>SAE</td>
<td>10,503</td>
<td>2007–2013</td>
<td>98.00%</td>
</tr>
<tr>
<td>Soui et al. (2020)</td>
<td>Ant colony optimization, LR, NN, RBF, RF</td>
<td>Ant colony optimization</td>
<td>250</td>
<td>N.A.</td>
<td>100.00%</td>
</tr>
<tr>
<td>Uthayakumar et al. (2020)</td>
<td>Ant colony optimization</td>
<td>SVM</td>
<td>1,957</td>
<td>2010–2016</td>
<td>83.60%</td>
</tr>
<tr>
<td>Sehgal et al. (2021)</td>
<td>LR, NN, SVM</td>
<td>SVM</td>
<td>1,957</td>
<td>2010–2016</td>
<td>83.60%</td>
</tr>
</tbody>
</table>

Source(s): Table created by author

Table 1. Overview of all included bankruptcy prediction studies

AB = AdaBoost; BN = Bayesian Network; CNN = Convolutional Neural Network; DT = Decision Tree; LD = Linear Discriminant Analysis; LR = Logistic Regression; LSTM = Long Short-Term Memory; MDA = Multivariate Discriminant Analysis; NB = Naïve Bayes; NN = Neural Network; RBF = Radial Basis Function; RF = Random Forest; SAE = Stacked Auto Encoder; SPSO = Switching Particle Swarm Optimization; SVM = Support Vector Machine
4.2 Financial analysis
The financial analysis deals with examining quantitative and qualitative accounting data to determine companies’ current and future economic situations. For this purpose, the annual financial statements are analyzed by internal and external analysts, who conduct their analysis for various (personal) motivations (Penman, 2013). In addition to the investor deciding whether to invest/disinvest and the bank deciding whether to grant a loan, a growing number of NGOs are also analyzing financial statements in order, for example, to conclude the sustainability of companies (Sisaye, 2021). Apart from investors, banks and NGOs, suppliers also have an interest in reliable financial analyses. On the one hand, to better assess the probability of the customer fulfilling the contract, and on the other hand, to have a planning basis for the development of the business relationship in the coming years. Therefore, automated financial analyses would help many different stakeholders to gain better insights into companies.

Forecasting future costs is a fundamental component of financial analysis, as this directly impacts a company’s future net assets, financial position and result of operations. An early study dealing with forecasting future costs was conducted by Boussabaine and Kaka (1998). In analogy to the studies on bankruptcy prediction, neural networks were used for predicting future costs. The actual and predicted cost curves of construction projects show only a little difference. The superiority of neural networks over other methods for predicting construction costs was also confirmed in the study by Karaca et al. (2020). Another study by Kuzey et al. (2019) identified future factors influencing cost system functionality using ML models. Especially the process of cost data management (i.e. collection, storage and use) has been identified as a key driver for an efficient cost system.

Besides forecasting companies’ future costs, financial analyses are often carried out to predict the future shareholder wealth (Machuga et al., 2002). The forecasting of shareholder wealth and stock returns has been critically discussed in academia for decades. Accordingly, there are already a large number of studies that use conventional statistical methods to evaluate the importance of preselected financial ratios for forecasting shareholder wealth and stock returns (Asquith and Mullins, 1983; Lewellen, 2004; Fischer and Lehner, 2021). Barnes and Lee (2007) were the first who investigated AI’s potential for analyzing which financial ratios are the main drivers of future shareholder wealth. Their forecasting model was based on neural networks, which performed best when the five input parameters – return on capital, share price, economic value-added, earnings per share and revenue – were used. Another study by Creamer and Freund (2010) uses boosting and tree-based algorithms to predict whether a company is over- or underpriced in the capital market and identifies non-linear relationships in accounting data. As a measure of the company’s capital market evaluation, Tobin’s q was applied. The financial ratio is calculated as a company’s market value divided by its assets’ replacement cost. Their results indicate that the random forest can predict most precisely the company’s Tobin’s q. The shareholder wealth also depends on which stage of the life cycle the company is currently in. The study by Lee et al. (2021) concludes that support vector machines can be used to predict the life cycle phase for the majority of companies.

Another essential part of evaluating companies’ future performances is financial liquidity. The study by Cheng and Roy (2011) analyzes to what extent AI-based forecasting methods help predict future cash flows. They found that a hybrid model consisting of support vector machines, fuzzy logic models and fast messy genetic algorithms provides reliable forecasts and performs better than single prediction models. An important part of cash flow management is also the forecast of outstanding invoices. The study by Bahrami et al. (2020) finds that the logistic regression outperforms support vector machines or the OneR algorithm in predicting customer payment dates. Besides delayed payments, payment defaults can also significantly impact the company that has been expecting the receivables. Payment defaults are also directly related to financial accounting, as IFRS 9 requires the use of the expected
credit loss model for the measurement of accounts receivable. The study by Martinelli et al. (2020) evaluates the extent to which payment defaults can be predicted ex ante. They show that the random forest is more accurate than neural networks or naïve Bayes and that almost all payment defaults can be predicted correctly. In another study by Sariannidis et al. (2020), support vector machines could predict payment defaults even slightly better than the random forest.

Besides cash flow, the forecast of future revenues is also a key indicator of companies’ developments. The study by Sai Vineeth et al. (2020) compares multiple ML models to identify the most accurate for predicting future sales. Unlike previous studies, ridge regression provided more accurate predictions than support vector machines or random forests. A different approach was used in the study by Kumar et al. (2021) to anticipate future revenues. For capturing competitive dynamic effects, they applied Bayesian generalized additive models. The results show that the forecast accuracy can be increased by up to 10% if market prices (including competitors’ prices) are considered. However, it should be criticized here that no alternative algorithms were used. So far, the previous studies have excluded the human factor. For the first time, the study by Yang et al. (2020) proposes a framework for how humans can be integrated into the forecasting process when the prediction is conducted using AI. In their study, the forecasting process is considered as a cycle in which humans are involved, from exploring the data to configuring the forecast models. The prediction models are tested iteratively until satisfactory results can be achieved. The main challenges are seen as ensuring that people have sufficient knowledge of the data and the prediction models. Without this knowledge, optimal results cannot be generated according to the current state of technology. A further study by Zhai et al. (2021) defines the requirements for applying AI-based forecasting models in accounting. According to their study, all sources of information must be available, and there must be clearly defined business processes. Furthermore, the input data must have a logical relationship to the desired output, every process has to be linked with an event timestamp, and the data must have a certain scale to ensure an efficient implementation.

In addition to cash flow and revenues, earnings are another important indicator of companies’ success. Financial analysts, in particular, pay great attention to companies’ earnings (Brown et al., 2015). The study by Shen (2012) evaluates the accuracy of neural networks for predicting earnings. Their main finding is that the more historical data is used as an input feature for training purposes, the better patterns can be identified for present and future earnings. The applicability of logistic regression for earnings forecasts was confirmed in a study by Baranes and Palas (2019). Their results indicate that a stepwise multivariate logistic regression can provide more accurate earnings forecasts than support vector machines. Another study by Quanyu et al. (2021) used long short-term memory neural networks for predicting future earnings. Their results show that the average prediction accuracy was substantially higher than those of financial analysts. However, for both mentioned studies (Shen, 2012; Quanyu et al., 2021), it can be criticized that no further forecasting models were considered for evaluation. Table 2 summarizes the studies mentioned above dealing with financial analysis in terms of the evaluated models, sample sizes and accuracies. If multiple data sets were analyzed, the data set with the highest accuracy is presented. Table 2 demonstrates that already few studies exist that predict cash flow, revenues and earnings. However, there is a lack of studies that aim to predict the total value of a company or at least the value of specific assets of companies.

4.3 Frauds and errors
As already mentioned, errors and frauds can cause enormous economic damage. According to IAS 8.5, errors are usually omissions or misstatements resulting from non-application or
misapplication of accounting standards. In contrast, fraudulent activities are intentional distortions to present the company’s financial situation usually better than it is (Rezaee, 2005). In practice, however, the distinction between fraud and error is often difficult, as a false statement may be due to intent or negligence of the reporting entity (Hung et al., 2017).

Investors and capital providers are particularly affected by frauds and errors, as they lose the capital they have invested due to the subordinate nature of their claims. Forecasting future frauds and errors of companies would help investors avoid misinvestments. In general, there are several techniques to detect frauds and errors in data sets. These range from traditional statistical methods to artificial immune systems, machine learning and

<table>
<thead>
<tr>
<th>Authors</th>
<th>Evaluated models</th>
<th>Best model</th>
<th>Forecast purpose</th>
<th>Sample size</th>
<th>Time</th>
<th>Evaluation metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Balance sheet variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creamer/Freund (2010)</td>
<td>AB, DT; LR; RF</td>
<td>RF</td>
<td>Company’s valuation</td>
<td>500</td>
<td>1992–2004</td>
<td>Error rate: 11.50%</td>
</tr>
<tr>
<td><strong>Profit and loss variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barnes/Lee (2007)</td>
<td>NN</td>
<td>NN</td>
<td>Shareholder Wealth</td>
<td>709</td>
<td>N.A.</td>
<td>R2: 0.956</td>
</tr>
<tr>
<td>Shen (2012)</td>
<td>NN</td>
<td>NN</td>
<td>Earnings</td>
<td>716</td>
<td>2005–2010</td>
<td>R2: 0.7211</td>
</tr>
<tr>
<td>Baranes/Palas (2019)</td>
<td>LR, SVM</td>
<td>LR</td>
<td>Earnings</td>
<td>3,877</td>
<td>2012–2017</td>
<td>Acc: 68.1%</td>
</tr>
<tr>
<td>Sai Vineeth et al. (2020)</td>
<td>GB, RF, RR, SVM</td>
<td>RR</td>
<td>Revenues</td>
<td>260</td>
<td>2014–2019</td>
<td>MAE: 0.058</td>
</tr>
<tr>
<td>Kumar et al. (2021)</td>
<td>BI</td>
<td>BI</td>
<td>Revenues</td>
<td>N.A.</td>
<td>N.A.</td>
<td>MAD: 6.79</td>
</tr>
<tr>
<td>Quianyu et al. (2021)</td>
<td>LSTM</td>
<td>LSTM</td>
<td>Earnings</td>
<td>3,865</td>
<td>2005–2014</td>
<td>Error rate: 11.39%</td>
</tr>
<tr>
<td><strong>Cash flow variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boussabaine/Kaka (1998)</td>
<td>NN</td>
<td>NN</td>
<td>Costs</td>
<td>N.A.</td>
<td>N.A.</td>
<td>RMSE: 0.001</td>
</tr>
<tr>
<td>Kuzey et al. (2019)</td>
<td>DT, LR, SVM</td>
<td>DT</td>
<td>Costs</td>
<td>565</td>
<td>N.A.</td>
<td>Acc: 91.5%</td>
</tr>
<tr>
<td>Bahrami et al. (2020)</td>
<td>LR, OneR, SVM</td>
<td>LR</td>
<td>Cash flow</td>
<td>1,659,083</td>
<td>2014</td>
<td>Acc: 97.39%</td>
</tr>
<tr>
<td>Karaca et al. (2020)</td>
<td>MLR, NN</td>
<td>NN</td>
<td>Costs</td>
<td>996</td>
<td>2006–2015</td>
<td>MAPE: 23.33%</td>
</tr>
<tr>
<td>Martinelli et al. (2020)</td>
<td>NB, NN, RF</td>
<td>RF</td>
<td>Cash flow</td>
<td>1,056,320</td>
<td>1995–1996</td>
<td>Acc: 99.8%</td>
</tr>
<tr>
<td>Sariannidis et al. (2020)</td>
<td>DT, LR, NB, NN, RF, SVM</td>
<td>SVM</td>
<td>Cash flow</td>
<td>30,000</td>
<td>2005</td>
<td>Acc: 82.21%</td>
</tr>
<tr>
<td><strong>Other variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee et al. (2021)</td>
<td>DT, NN, SVM</td>
<td>SVM</td>
<td>Company’s life cycle</td>
<td>4,498</td>
<td>2011–2018</td>
<td>Acc: 68.48%</td>
</tr>
</tbody>
</table>

Table 2. Overview of all included prediction studies dealing with financial analysis

AB = AdaBoost; Acc = Accuracy; BI = Bayesian Inference; DT = Decision Tree; FSS = Fuzzy Support System; GB = Gradient Boosting; LR = Logistic Regression; LSTM = Long Short-Term Memory; MAE = Mean Absolute Error; MAD = Mean Absolute Deviation; MAPE = Mean Absolute Percentage Error; MLR = Multiple Linear Regression; NB = Naive Bayes; RF = Random Forest; RMSE = Root-Mean-Square Error; RR = Ridge Regression; SVM = Support Vector Machine

Source(s): Table created by author
meta-learning methods (Donning et al., 2019). All of the articles discussed below use machine learning algorithms that are trained by supervised learning.

The study by Rahul et al. (2018) compares the suitability of supervised learning methods with unsupervised learning methods for financial fraud detection. The results show that the Gaussian distribution, which belongs to unsupervised learning, has a higher prediction accuracy than the random forest or other boosting methods, such as AdaBoost or XGBoost. In addition, the predictive power of the DBSCAN algorithm is evaluated in the study of Tatusch et al. (2020). The algorithm is based on a clustering method and is therefore categorized as unsupervised learning. By using only two or three input features, the algorithm can predict more than half of the correct errors.

Furthermore, the question arises of which input variables are relevant for predicting frauds and errors. Besides using only financial ratios as input features, the study by Bao et al. (2020) shows that raw accounting numbers can also be considered for training prediction models. Another study by Brown et al. (2020) finds that text-based measures of the thematic content of financial statements are useful for detecting misreporting. Prediction models that detect and quantify the thematic content of financial statements outperformed models that exclusively use financial or unprocessed textual data. Their qualitative content analysis was conducted by applying the bag-of-words algorithm utilizing the distribution of words across documents to classify and quantify topics without requiring predefined word lists. The increased data quality by combining different data sets was also demonstrated in the study by Bertomeu et al. (2021). They found that accounting variables are more informative as input features for predicting material misstatements if used together with audit and market variables. Another study by Rahman et al. (2021) showed that incorporating ownership features could substantially improve forecast accuracy. In their study, the naïve Bayes algorithm provided the best predictions.

Besides choosing suitable algorithms and additional input features for predicting frauds and errors, it should be evaluated how algorithms can be incorporated into audit activities. The study of Sun (2019) shows the possibility of providing additional features to prediction models after the first prediction. Based on the initial forecast, the auditors can initiate further investigations and collect new variables. These variables can be used as new input features for recalculating the predictions. This process can be repeated as often as desired, allowing dynamic and flexible auditing procedures. Table 3 summarizes the research dealing with forecasting frauds and errors in terms of the evaluated models, sample sizes and accuracies. If multiple data sets were analyzed, the data set with the highest accuracy is shown.

5. Discussion and open research

After examining the prediction models used in previous studies and their accuracy on an application-specific basis, the following section analyzes whether similar prediction models are used across applications. To address this, Table 4 presents an overview of all prediction models applied in previous literature. In addition, Table 4 contains an assignment of which prediction model is suitable for supervised and unsupervised learning methods.

It becomes evident that support vector machines, logistic regression, neural networks and tree-based methods such as random forests and decision trees are often used for forecasting tasks in financial accounting. However, it is noticeable that many studies only consider limited algorithms on data sets that have not been investigated by other studies before. Thus, there is still little knowledge about the extent to which one algorithm is superior to another in terms of prediction accuracy. A possible way to address this problem would be to evaluate under which conditions (e.g. data structure, distribution and sample size) a forecast model performs better than others in terms of accuracy and robustness. Additionally, it can be observed that recent studies use different metrics for evaluating similar prediction tasks.
### Table 3.
Overview of prediction studies dealing with fraud and error detection

<table>
<thead>
<tr>
<th>Authors</th>
<th>Evaluated models</th>
<th>Best model</th>
<th>Forecast purpose</th>
<th>Sample size</th>
<th>Period</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rahul et al. (2018)</td>
<td>AB, GD, RF, XGBoost</td>
<td>GD</td>
<td>Frauds</td>
<td>1,239</td>
<td>2005–2015</td>
<td>71.68%</td>
</tr>
<tr>
<td>Bao et al. (2020)</td>
<td>LR, RB, SVM</td>
<td>RB</td>
<td>Frauds</td>
<td>206,026</td>
<td>1979–2014</td>
<td>75.3%</td>
</tr>
<tr>
<td>Brown et al. (2020)</td>
<td>LDA</td>
<td>LDA</td>
<td>Errors</td>
<td>131,528</td>
<td>1994–2012</td>
<td>74.2%</td>
</tr>
<tr>
<td>Bertomeu et al. (2021)</td>
<td>DT, GB, LD, LR, NB, RB, RF, SVM</td>
<td>RB</td>
<td>Errors</td>
<td>54,354</td>
<td>2001–2014</td>
<td>76.3%</td>
</tr>
<tr>
<td>Rahman et al. (2021)</td>
<td>DT, NB, NN, RF</td>
<td>NB</td>
<td>Errors</td>
<td>2,262</td>
<td>2000–2007</td>
<td>63.6%</td>
</tr>
</tbody>
</table>

Table 4. Overview of all prediction models included in our final sample and their possible forecasting tasks

<table>
<thead>
<tr>
<th>Models</th>
<th>Quantity</th>
<th>Supervised learning</th>
<th>Unsupervised learning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Classification</td>
<td>Regression</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>18</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>17</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Neural Network</td>
<td>15</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Random Forest</td>
<td>11</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>10</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>6</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AdaBoost</td>
<td>3</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Multivariate Discriminant Analysis</td>
<td>3</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Bayesian Network</td>
<td>2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Gradient Boosting</td>
<td>2</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Linear Discriminant Analysis</td>
<td>2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Long Short-Term Memory</td>
<td>2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Nearest Neighbour</td>
<td>2</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>RUSBoost</td>
<td>2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Stacked Auto Encoder</td>
<td>2</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>XGBoost</td>
<td>2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>ZeroR</td>
<td>2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Ant Colony Optimization</td>
<td>1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bayesian Inference</td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Convolutional Neural Network</td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Fuzzy Support System</td>
<td>1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Gaussian Process</td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Gaussian Distribution Model</td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Kohonen Network</td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Linear Dirichlet Allocation</td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Multiple Linear Regression</td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Radial Basis Function</td>
<td>1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ridge Regression</td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Switching Particle Swarm Optimization</td>
<td>1</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

**Source(s):** Table created by author
Depending on which evaluation metric is used, the suitability of certain prediction models might differ. Therefore, more generalizable knowledge about how AI models should be constructed and used is needed. To address this, future researchers should create design requirements and design principles that can be adapted to various common prediction problems practitioners and researchers face.

Open data is another concept that might help to support and advance research of AI-based forecasting in financial accounting. According to the Open Knowledge Foundation, the term open knowledge refers to “any content, information or data that people are free to use, re-use and redistribute […] without any legal, technological or social restriction” (Open Knowledge Foundation, 2022). Open data repositories have numerous advantages, including easier access to data, lower costs, easier knowledge transfer to industry and the reuse of data (for a detailed overview of benefits, see Kitchin et al. (2015)). Data that are freely and easily accessible allow for a better comparability of different AI algorithms that were applied on the same data set by different research groups. Therefore, the availability of open accounting data repositories would be highly beneficial for future research in the area of AI-based forecasting to investigate which AI algorithms are most suitable for a given task.

A methodology common in information systems (IS) that might be highly suitable for closing this research gap is design science research (DSR). DSR aims to create new solutions or artefacts for real-world problems (vom Brocke and Maedche, 2019). The results of DSR can be sociotechnical artefacts and design knowledge, both to investigate why a certain artefact enhances a specific application context (vom Brocke et al., 2020; Gregor and Hevner, 2013). Unfortunately, none of the studies mentioned above uses a systematic DSR approach to develop their respective artefacts. Instead, existing research focuses mainly on technical aspects of the underlying AI systems. Researchers have therefore generated knowledge about certain IT artefacts, how they work, and their accuracy. However, the knowledge that also considers social or organizational aspects is still missing. Thus, in addition to certain IT artefacts, also IS artefacts should be considered (Lee et al., 2015). Researchers can create valuable and new contributions by developing and evaluating IS artefacts. These artefacts allow addressing implementation issues not covered by computational science or accounting research.

A key implementation issue for AI predictions in accounting is also the human factor. First, the acceptance of employees to use these models is important to consider. According to current surveys, the acceptance of employees to use AI applications depends on the clarity of user and AI roles as well as trust in AI systems (Choi, 2021). However, to the best of our knowledge, no studies have yet been conducted investigating employee acceptance of AI-based forecasts by accounting professionals or examining their role and range of tasks within the forecasting cycle. This research gap can be addressed by applying a technology acceptance model, such as the unified theory of acceptance and use of technology (UTAUT), to explain whether and how employees would use AI-based forecasts in accounting (Venkatesh et al., 2016). UTAUT is also compatible with DSR, as the theory provides a knowledge base to develop IS artefacts that aim to increase employee acceptance of using AI models. In this context, future research can develop different IS artefacts and test them iteratively by measuring the extent to which employees’ acceptance is increased.

Another aspect that might also increase the acceptance is the high explainability of the AI methods used. The research field of explainable AI is of growing importance and involves research directions used in other studies (Bauer et al., 2021). For example, when employees must make important business decisions based on forecasts, the comprehensibility of the forecast is an important factor. However, the decision rules of complex forecasting models are often difficult to understand, especially for humans without specific technical knowledge. To address this problem, IS artefacts can be designed and evaluated that aim to increase the comprehensibility of AI models. Besides the acceptance of employees, future research should
also investigate at which stages employees can participate in the forecasting process. Since the human factor plays an essential role in implementation issues (Grover and Lyytinen, 2015), it must be examined how the interaction between employees and AI models can work efficiently. Furthermore, it is of high relevance to answer the question what tasks are performed by accountants, and what is done by the AI systems. We can expect that AI systems will be used for accounting tasks of increasing complexity in the future (Leitner-Hanetseder et al., 2021; Skrbić and Laughland-Booy, 2019). Due to these developments, AI will be capable of performing more and more tasks that required human intelligence and involvement before (Huttunen et al., 2019; Leitner-Hanetseder et al., 2021). Therefore, the profession and tasks of the human accountant will be subject to continuous development in the future. In this context, it is also important to explore the ethical implications of using AI in more detail (for recent overviews, see, e.g. Munoko et al. (2020) or Lehner et al. (2022)).

Furthermore, the successful implementation of AI-based forecasts into accounting information systems requires concrete information regarding their maintenance. It must be criticized that previous prediction studies have not evaluated after which period the forecast models have to be retrained to continue making reliable predictions. This could also be a reason why AI forecasts are still often pilot projects that are unsuccessful in the long term and are therefore discontinued. This issue can be addressed by collecting long-term observations for similar prediction problems to provide recommendations when prediction models must be retrained.

Moreover, it is noteworthy that the current forecasting application fields in financial accounting focus on a few use cases. Especially in financial analysis, there are several potential use cases that have not yet been investigated in detail. Given that most AI-based forecasting studies concentrate on profit and loss or cash flow, the forecasting of future balance sheet figures is a major research gap. In addition, the utility of other technologies such as robotic process automation for AI-based automated forecasting purposes has been inadequately investigated (Onyshchenko et al., 2022).

Summarizing the open research agenda, the research field of AI-based forecasting in accounting can benefit significantly from further insights from future investigations, mainly dealing with implementation issues in accounting. Previous studies proposed mainly technical solutions that are isolated from each other. Thus, future research could merge the results of previous research to create generalizable knowledge that can serve as a solution for a group of problems dealing with AI-based forecasting in accounting. This would help future researchers and practitioners to facilitate and shorten the process of finding and implementing the most accurate prediction model.

6. Conclusion
In recent years, AI has made it possible to search for patterns of increasing complexity. The patterns identified can be used for predicting companies’ developments. Especially in accounting, it can be observed that prediction studies increasingly contain larger sample sizes over time and that the models produce more accurate forecasts. In this paper, we investigated the current research status in the field of AI-based predictions in financial accounting and provided the first systematic literature review in this emerging field of research. Our first research goal was to present an overview of how AI-based technologies and approaches are used for forecasting tasks in financial accounting. Furthermore, we aimed to outline a future research agenda from a holistic perspective. To answer these research questions, we conducted a systematic literature review of several scientific databases to identify relevant research and publications.

Our findings indicate that, so far, there are three main application fields for AI-based forecasts in financial accounting. The areas of application range from bankruptcy forecasts to financial analysis as well as fraud and error detection. Especially existing and future
investors can benefit from these predictions, as knowledge about future business developments helps to avoid bad investments. Previous studies show that support vector machines, neural networks, and random forests provide accurate and robust predictions for all three application areas.

However, there is little evidence of whether one prediction model provides significantly more accurate forecasts than others. Especially for financial analysis, it is difficult to derive generalizable knowledge, as different financial variables are predicted (e.g. revenues, cash flow and earnings). Future research needs to evaluate the predictive performance of models on different data sets with common properties to address this research question. Furthermore, most studies do not consider issues related to the implementation and maintenance of forecasting models. Thus, the human role in the forecasting process is mostly neglected by prior research. Future research can make a significant contribution to closing these research gaps and more closely link computational science and accounting research. By using DSR, IS artefacts can be developed and evaluated besides IT artefacts, holistically addressing the implementation of prediction models in organizations.

References


Corresponding author
Marko Kureljusic can be contacted at: marko.kureljusic@uni-due.de