ISSN 1870-4069

A Systematic Mapping Study of Artificial Intelligence in Software Requirements

Alfonso Robles Aguilar, Jorge Octavio Ocharán Hernández,
Ángel Juan Sánchez García

Universidad Veracruzana,
Facultad de Estadística e Informática,
Mexico.

alfonsoroag8@gmail.com,
{jocharan, angesanchez}@uv.mx

Abstract. Artificial Intelligence techniques are used in a wide variety of domains. The application of Artificial Intelligence (AI) in Software Engineering has assisted activities that are time-consuming and resource-intensive such as those related to software requirements. This paper focuses on the use of AI techniques in Requirements Engineering (RE). The RE activities are critical in the life cycle of the software since they are the foundation of further phases. This paper presents the results of a Systematic Mapping Study that reports the extensive use of AI techniques in different RE activities, including recent trends, results, and future work.

Keywords: Requirements Engineering, Artificial Intelligence, Systematic Mapping Study, Software Development.

1 Introduction

In recent years, the interest in the application of Artificial Intelligence (AI) techniques, algorithms, and methods in Software Engineering (SE) has increased, seeking to address the problems and challenges that arise in the creation of software systems [1]. AI represents an opportunity to automate processes in the software industry. As a result, the successful use of AI techniques in almost all areas of SE has been reported, highlighting search, optimization, classification, learning, and prediction algorithms [1]. Within the SE, the closest phase to the user is the Requirements Analysis [2], that is part of a whole discipline called Requirements Engineering (RE), which in addition to the analysis, covers the elicitation, specification, and verification of the Software Requirements (SR) [3]. A SR is a capability that must be met or possessed by a software system to satisfy a contract, standard, specification, or other documents [4]. This definition allows us to glimpse the importance that RE has in SE since they reflect the user's needs for the system.

The aim of this paper is twofold. Firstly, to make a succinct presentation of the state the art in the use of AI techniques in RE activities in such a way Requirements Engineers could consult a compilation of the existing techniques to find a solution to their problems or improve their processes. Secondly, to identify research gaps in the use and optimization of AI techniques.

This paper is organized as follows. Section II presents previous works on the use of AI in RE activities. Section III presents the systematic review method used in this research. Section IV discusses the results of this study, and Section V presents the contribution of this paper and our conclusions.

2 Related Work

The interest to know methodologies to support RE has increased, such as the review conducted by Dermeval *et al*., where they report the application of ontologies in RE as well as how ontologies support RE [5].

Due to Harman's report of increased interest in using AI in SE [1], we can assume that there is also research in this area, and AI techniques have been applied in RE seeking to improve the processes and results of their activities. As evidence of the interest in this area, five reviews related to this Systematic Mapping Study (SMS) were found in preliminary research.

Sorte *et al*. [6] reviewed the state of the art of AI use throughout the software development life cycle. This implies that no area is delved into, such as RE, with only four papers are reported, which gives us the opportunity to explore deeper this topic.

Águila and del Sagrado [2] focus on finding how and where Bayesian Networks (BN) are applied, an area of AI [7], in the RE. This paper suggests that BN has not been applied in all areas of the RE, but that there is synergistic cooperation between these two, as they report cases in which some RE activities have been improved using BN. Since this paper is from 2016, we can assume that, due to the reported synergy, the investigation has been continued in the use of BN in RE, and there are works from more recent years that were not reported. This also justifies the creation of our paper, in addition to the fact that we wanted to cover more areas of AI and not just BN.

Gramajo *et al*. [8] and Iqbal *et al*. [9] expose papers where Machine Learning (ML) techniques are used, another area of AI [7], in the RE. Both reviews present the use of ML algorithms exclusive int RE, such as Support Vector Machine (SVM), Naïve Bayes, and Neural Networks. The purpose of our mapping, unlike these two, is to report more than one area.

Finally, Haq *et al*. [10] have a review similar to this one, in which they explore the uses of AI in RE, with the difference that they classified RE in phases, this study that focuses on RE activities. Furthermore, except for the first and second International Workshop on Artificial Intelligence for Requirements Engineering, the sources from which the studies were obtained are not explicitly specified nor the search string used in the electronic databases.

As can be seen from the number of reviews, there is considerable interest in the use of AI in RE, so a more updated version with scope for more AI areas may be more useful for researchers in the area.

Table 1. Keywords.

Concept	Keywords
Requirements Engineering	software requirements engineering, requirements engineering, software requirements analysis, software requirements specification, software requirements verification
Artificial Intelligence	Artificial Intelligence

3 Research Method

This SMS was conducted according to the guidelines proposed by Kitchenham and Charters [11]. They include the specification of the research questions, primary study search and selection, and data extraction and data synthesis processes.

3.1 Planning

Research Questions. The research questions (RQs) relating to this SMS are:

- RQ1: What are the AI techniques used in RE?
- RQ2: What RE activities have been addressed with these techniques?

The motivation of the RQ1 is to identify the use of AI techniques that have been applied in RE. Regarding the reason for RQ2 is to discover in which specific RE activities, AI techniques have been applied.

Search strategy. The search strategy was an automated search, looking for the primary studies among the results that electronic databases show from a search string.

Keywords. These words (Table 1) were selected based on the Research Questions. Regarding RQ1, the most important word that was found was "Artificial Intelligence," which resulted in a Keyword for the search. Regarding RQ2, "Requirements Engineering" was identified, and as Keywords, all phases of the RE were selected to cover every aspect of the process in the search.

Search string. After trying different strings using the keywords and analyzing the results they showed, the following search string was formulated, in order to cover studies that apply RE or at least some of its phases:

("software requirements engineering" OR "requirements engineering" OR "software requirements analysis" OR "software requirements specification" OR "software requirements verification") AND ("artificial intelligence")

It should be clarified that one of the limitations of our study is that by wanting to cover all AI, it is likely that the search chain omits papers that are valuable for our mapping but if the paper didn't use the concept AI, the chain could not show them.

Table 2. Inclusion criteria.

ID	Description
IC1	They were published between 2015 and 2019.
IC2	Written in English.
IC3	The title, abstract, or keywords suggests that the paper answers at least one research question.
IC4	Answer at least one research question.

Table 3. Eclusion criteria.

ID	Description
EC1	Reports the requirements of some products other than software.
EC2	It is not a journal or proceedings paper.
EC3	Reports requirements for an AI system but does not use AI in the requirements process.
EC4	Duplicate papers (same research in different databases).

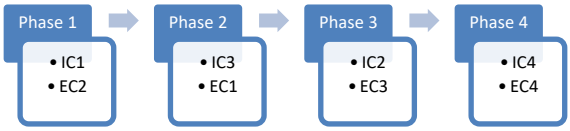

Fig. 1. Primary studies selection process.

Electronic database resources. The sources used for the search and selection of primary studies are four electronic databases: ScienceDirect, SpringerLink, IEEE Xplore, and ACM Digital Library. These four sources of information were chosen because they contain papers from the area of computer science, SE, and related disciplines, following the path marked by the Keywords to obtain the necessary papers in this area. Furthermore, when published in these electronic databases, the papers went through an arbitration process by experts in the field.

Primary study selection criteria. Table 2 and Table 3 show the used criteria to determine if a study was considered a Primary Study (PS) or the paper did not meet what was necessary to be included.

Primary study selection procedure. The phases in which these selection criteria were applied to choose the PS are the ones shown in Fig. 1. When doubts were raised regarding the inclusion of the papers, the second author (AI expert) and the third author (SR expert) decided whether the paper should be a primary study or not.

Data extraction. Data were extracted using the extraction template shown in Table 4. We extracted two kinds of data from each one of the primary studies selected, that were considered necessary in order to answer the research questions. As recommended by the method, the first five data extracted correspond to publication details. The sixth and seventh data are directly related to RQ1 and RQ2 and correspond to context information; these data can answer both questions.

Table 4. Data extraction template.

Kind of data	Data extracted
Publication details	Title
	Authors
	Year
	Source
	Author keywords
Context	AI technique
	RE activity

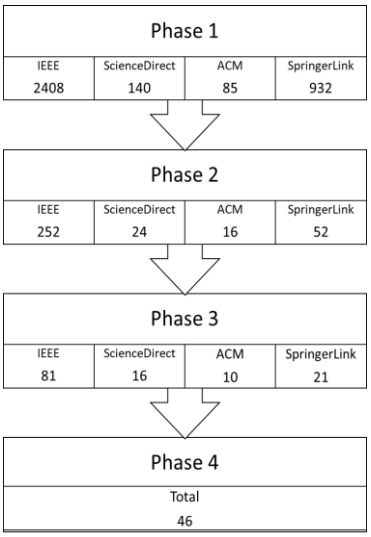

Fig. 2. Primary studies selection phases.

Data synthesis. For data synthesis, the Thematic Synthesis strategy proposed by Cruzes and Dyba [12] was followed for the analysis of qualitative data. This is because this is a method that has been found useful in qualitative data analysis in the SE area reviews.

3.2 Conducting

As shown in Fig. 2, the string applied in digital libraries yielded a total of 3,565 documents, applying the publication date, and study type filters (Phase 1). After Phase 2, a total of 344 papers made it to Phase 3, and in this phase, 128 papers met the selection criteria (IC2, EC3). Finally, 46 papers fulfilled all the selection criteria, and these were selected as PS.

The process for carrying out this review is like many other reviews in the area of Software Engineering since it used methodologies already used by experts in the field. In order to achieve this, the previously mentioned guidelines were followed.

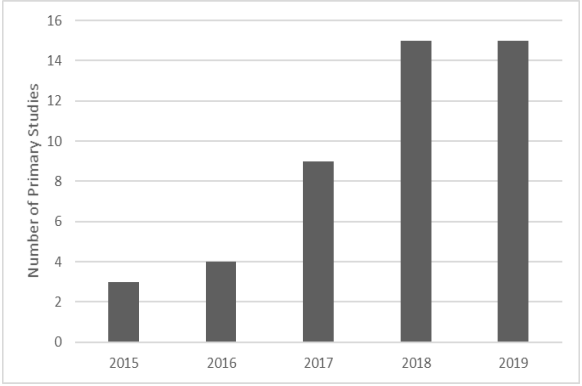

Fig. 3. Number of papers per year publication.

4 Results

Forty-six studies were selected, where 102 cases of the use of some AI technique in some RE activity were reported since some papers reported using two or more techniques in one or more activities. Regarding the authors, 141 different names were identified, of which only ten were present in more than one paper (precisely two papers each), these are Davide Dell'Anna, Fabiano Dalpiaz, Zahra Shakeri Hossein Abad, Didar Zowghi, Martin Glinz, Guenther Ruhe, Ralph Samer, Muesluem Atas, Alexander Felfernig, and Atsushi Ohnishi.

It can be seen in Fig. 4 a definite growing interest in the use of AI in RE in recent years. Having three papers published in 2015 to 15 in 2018. We can observe a clear trend since publications have quintupled in just two years and maintaining interest in 2019 with 15 publications equally to the previous year.

In order to answer the research questions, the thematic synthesis process shown in Fig. 5 was followed, and 37 themes were obtained from the data that answer the research questions. The relationship among these themes is shown in Fig. 6. In respect of techniques, these were organized into groups of algorithms. Of the 46 papers, 102 reports were obtained, which were grouped into 20 higher-order themes. To achieve this, techniques that belong to the same family, such as Naïve Bayes Multinomial and Naïve Bayes, were put in the same theme.

To cite another example, neural networks such as the Convolutional Neural Network and the Recurrent Neural Network, as well as the rest of the Neural Networks, were placed in the same theme. Regarding RE activities, were grouped in 17 higher-order themes of what best described the work performed by the technique in the RE process.

For example, the cases in which requirements were classified were grouped into one (Requirements Classification), whether they were classification of non-functional, functional, security requirements, or any other way of classifying them. Another example was the theme classified as "Requirements Extraction" Activities that extracted requirements from a document such as AppStore and Twitter reviews were located here.

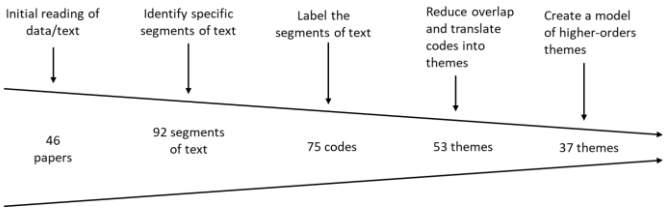

Fig. 4. Thematic synthesis process.

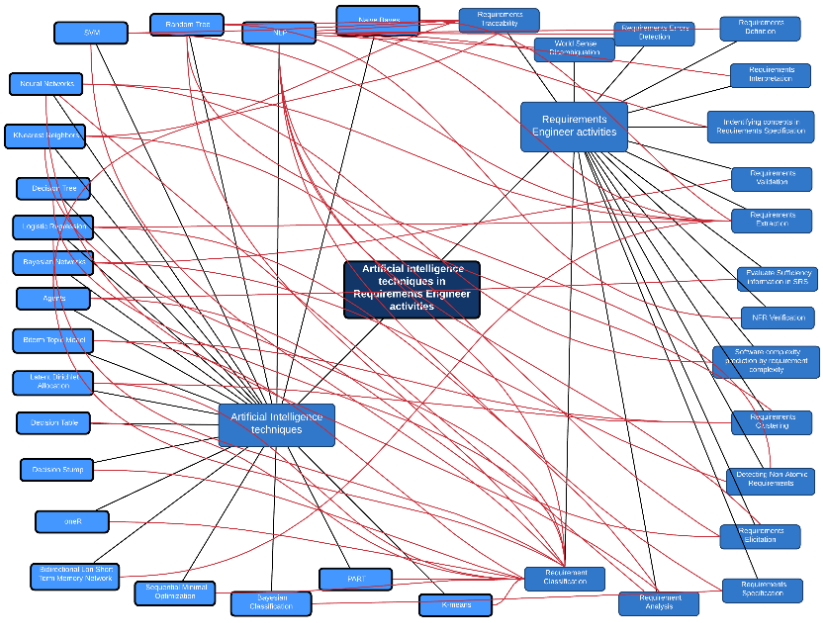

Fig. 6. Thematic map result.

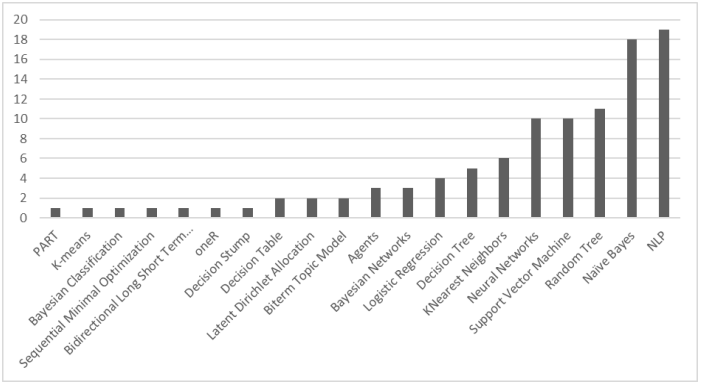

Fig. 7. Number of reports per AI technique.

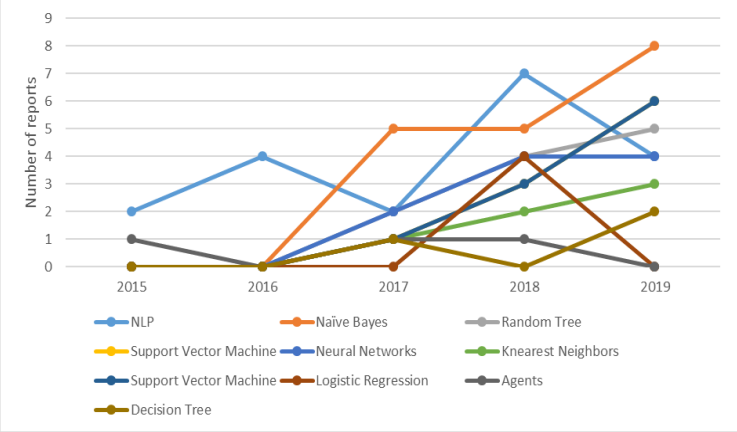

Fig. 8. Number of main AI techniques reports per year.

In response to **RQ1**, we found that the use of 20 different AI techniques in RE activities has been reported. We can see the extensive use of Natural Language Processing (NLP) since it accounted for 19 reports, representing 18.62% of the total. Also, Naïve Bayes reports the use of 18 cases in RE activities, representing 17.64% of cases. Random Tree reported 11 cases and tied with ten reports each, we found that SVM and Neural Networks also have records of being used in RE.

The last five techniques mentioned sum up 68 reports, which represent 66.68% of the total, suggesting that there is a remarkable concentration in the use of these five techniques since the remaining 15 techniques only account for a third of the total. The remaining can be found in Fig. 7.

As shown in Fig. 8, in the last three years there has been an increasing trend in the use of NLP, K-Nearest Neighbors, Naïve Bayes, SVM, Neural Networks, and Random Tree. Logistic Regression had a remarkable use in 2018 but had no record in any other year. It can be seen a clear trend in the use of Naïve Bayes.

In order to answer **RQ2**, a total of 17 different activities (Fig. 9) were found that are reported to have been carried out by IA. As we can notice, there is an extensive study in the Requirements Classification. Of the 102 cases reported, 59 sought ways to classify requirements, representing 57.84% of the total, a significant amount considering that it is only one of 18 activities. Given the large amount of work on requirements classification, a requirement engineer will be able to find many suggestions to use. Regarding Requirements Analysis and Requirements Extraction, nine and eight reports were found, respectively, in which AI techniques were used.

In the table in Appendix A[1], the specific use that is given to AI techniques in RE activities can be seen. With it, we can identify the relations that exist between AI techniques and RE activities, as well as the references to all 46 papers selected as primary studies (Appendix B[2]). Appendix A and B are available on the Web.

In addition, we can see that Naïve Bayes has been used 16 times in Requirements Classification throughout 11 papers (this happened because some papers use more

[1] https://bit.ly/3kVoGHV
[2] https://bit.ly/2S93PEq

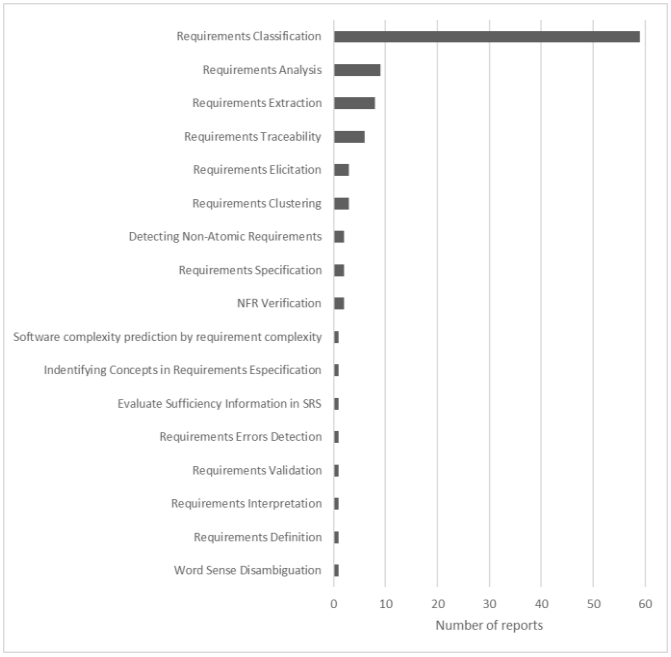

Fig. 9. Number of reports per RE activity.

than one type of Naïve Bayes). The following two areas with most reports (tied with eight reports each in seven papers) are the use of SVM and Random Trees, both in Requirements Classification. In this table, it can be seen the vast catalog of AI techniques that exist to classify requirements: 17 ways to classify AI requirements were found in this SMS. Only 3 of all the AI techniques in this work have not been applied for this activity.

Another point to note is that the Requirements Analysis, even being the second most reported activity, has only been addressed by two AI techniques: NLP and Neural Networks. Another noteworthy point is that the Requirements Extraction, although it only has eight reports, has tried to tackle with six different types of AI techniques. The table in Appendix A can be an excellent opportunity for researchers to get to know the least studied areas and have quick access to existing records.

5 Conclusions and Future Work

An SMS was conducted to answer the two research questions, applying a search string in four electronic databases where 46 studies were found that report of the use of AI techniques in RE. After analyzing those studies, a detailed summary of the AI techniques that have been used in the RE was drawn, where the use of 20 techniques in 17 different RE activities is reported. The most used techniques are NLP and Naïve Bayes. The most addressed activity is Requirements Classification with 59 reports. In them, it is noted that this activity has been carried out with 19 different AI techniques, being Naïve Bayes the most used with 16 reports. As future work, as this paper is part

of more extensive research work, we will seek to know the use of AI in other areas of Software Engineering, as well as obtain in more detail the data from this paper.

References

1. Harman, M.: The role of artificial intelligence in software engineering. In: Proceedings of the First International Workshop on Realizing AI Synergies in Software Engineering, pp. 1–6, IEEE Press (2012)
2. Del Águila, I.M., Del Sagrado, J.: Bayesian networks for enhancement of requirements engineering: a literature review. Requir. Eng., 21, pp. 461–480 (2016)
3. Bourque, P., Fairley, R.E.: SWEBOK: guide to the software engineering body of knowledge. IEEE Computer Society (2014)
4. ISO/IEC/IEEE: ISO/IEC/IEEE International Standard - Systems and software engineering-Vocabulary. ISO/IEC/IEEE 24765:2010(E), pp. 1–418 (2010)
5. Dermeval, D., Vilela, J., Bittencourt, I.I., Castro, J., Isotani, S., Brito, P., Silva, A.: Applications of ontologies in requirements engineering: a systematic review of the literature. Requir. Eng., 21, pp. 405–437 (2016)
6. Sorte, B., Joshi, P., Jagtap, V.: Use of artificial intelligence in software development life cycle: A state of the art review. Int. J. Technol. Manag., 03, pp. 2309–4893 (2015)
7. Russell, S., Norvig, P.: Artificial intelligence: A modern approach. Pearson (2016)
8. Gramajo, M.G., Ballejos, L., Ale, M.: Software requirements engineering through machine learning techniques: A literature review. In: IEEE Biennial Congress of Argentina (ARGENCON), pp. 1–7 (2018)
9. Iqbal, T., Elahidoost, P., Lúcio, L.: A bird's eye view on requirements engineering and machine learning. In: 25th Asia-Pacific Software Engineering Conference (APSEC), pp. 11–20 (2018)
10. Haq, B., Nadeem, M., Ali, I., Ali, K., Raza, M., Rehmanr, M.U.: Use of expert system in requirements engineering process a systematic literature review. In: 2019 UK/ China Emerging Technologies (UCET), pp. 1–5 (2019)
11. B. Kitchenham, S.C.: Guidelines for performing systematic literature reviews in software engineering (2007)
12. Cruzes, D.S., Dyba, T.: Recommended steps for thematic synthesis in software engineering. In: International Symposium on Empirical Software Engineering and Measurement, pp. 275–284 (2011)