IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 April 2024, accepted 26 April 2024, date of publication 30 April 2024, date of current version 7 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3395369

TOPICAL REVIEW

Artificial Intelligence for Quality of Life Study: A Systematic Literature Review

AYOUB JANNANI, NAWAL SAEL, AND FAOUZIA BENABBOU

Laboratory of Information Technology and Modeling, Faculty of Sciences Ben M'Sick, Hassan II University of Casablanca, Casablanca 20000, Morocco

Corresponding author: Ayoub Jannani (jannaniayoub@gmail.com)

ABSTRACT The concept of "quality of life" (QoL) encompasses all aspects of people's standard of living, including economic, social, or health-related factors, as well as their perceptions of their own lives. Although the growing application of Artificial Intelligence (AI) and Machine Learning (ML) techniques in processing and modeling the diverse datasets associated with these domains, there remains a significant challenge in addressing different issues in QoL area of research and in fully harnessing these technologies to improve QoL research outcomes. Despite the technological advancements, current research endeavors often overlook the complex, multifaceted nature of QoL study. This oversight results in fragmented insights and leaves significant areas underexplored. In this work, we conducted a systematic literature review (SLR) to investigate the contribution of AI to QoL studies. For this, we collected 68 research works published between 2008 and 2022. This review covers a range of research questions about the objectives and methods of studies on QoL, the sources and types of data utilized, and the advancements made through the application of natural language processing (NLP), ML, deep learning (DL), statistical models, and semantic approaches. The goal of this review is to tackle the prevalent ambiguity in QoL dimensions, synthesize the research findings, and highlight the contributions, advancements, and most innovative approaches in the field. Moreover, we identify gaps and limitations in the current literature and suggest potential areas for future research, aiming to inspire more cohesive and comprehensive approaches to studying QoL using AI and ML techniques.

INDEX TERMS Artificial intelligence, machine learning, deep learning, natural language processing, ontology, quality of life, well-being, happiness.

I. INTRODUCTION

Artificial intelligence technologies are used for a variety of tasks across a wide range of research domains, including medicine, transportation, industry 4.0, and the social sciences. Quality of life is a highly complex and multifaceted social science concept, and its assessment and analysis present significant challenges to researchers. While there is growing recognition of the importance of measuring and promoting wellbeing, there is still considerable debate about the most appropriate indicators to use [1], [2].

The United Nations Sustainable Development Goals (SDGs) aim to promote sustainable development and improve the well-being of individuals worldwide. Advances in technology, particularly in the field of artificial intelligence, offer new opportunities to identify and measure indicators of wellbeing in more efficient and accurate ways, which can be helpful in achieving multiple SDGs.

QoL studies aim to investigate all facets of people's living conditions as well as their perceptions of their own lives. This concept is related to different fields, including economics, health, and social sciences, and can be studied at different levels, such as country, city, region, or individual [1]. The terms "happiness," "well-being," and "quality of life" are used as synonyms in the literature [3].

Subjective well-being (SWB) is a broad concept related to several aspects of a person's well-being, including life satisfaction, positive and negative affect, and psychological well-being. The study of SWB is important not only for individuals but also for the whole society [4]. Policymakers

The associate editor coordinating the review of this manuscript and approving it for publication was Jolanta Mizera-Pietraszko.

around the globe are beginning to consider the use of SWB measures to monitor progress and evaluate public policies [5].

However, a fair assessment of the quality of life (QoL), especially on the country level, must take into account more than just people's sentiments and individual satisfaction, but also other factors, particularly objective measures such as the economic dimension [6], [7].

Hence, multiple international organizations have proposed the use of QoL indicators that take into account multiple elements of living standards in order to help standardize and measure well-being in countries and regions around the world. These indicators aim to provide a more comprehensive understanding of QoL by considering a range of factors that can impact country-level well-being, such as income, education, health, and other social and economic indicators [8]. The goal of these indicators is to help provide a more standardized way of measuring QoL across different countries and regions. These indicators include the world happiness index (WHI), the better life index (BLI), gross domestic product (GDP), and gross national happiness (GNH).

Additionally, health-related quality of life (HRQoL) is typically concerned with examining an individual's physical and mental well-being, and overall health status [9].

The QoL study demonstrated that individual well-being is closely tied to the well-being of communities and countries [10], [11]. The "quality of urban life" (QoUL) has gained increasing importance in studies of well-being at the city level due to its relevance in the planning and management of urban policies [12].

ISO 37120, proposed by the International Organization for Standardization, aims to promote sustainable development in cities through the assessment of the impact of smart city initiatives and the provision of a set of indicators to aid policymakers in improving city services and quality of life [13].

Recent research on quality of life has benefited from AI techniques thanks to their high performance in processing big data and extracting information. Specifically, ML and DL algorithms are used to construct models for analyzing, forecasting, and assessing well-being. Furthermore, research on quality of life can be advanced through the use of semantic tools during the stage of knowledge representation [14], which helps in the unification of QoL indicators and in storing their data in a uniform way.

Despite the fact that the topic of well-being has been widely covered in the literature, the studies carried out are mostly in the sociology, health, or psychology domains. To the best of our knowledge, there are only a few review articles that address the use of AI algorithms for studying the concept of QoL.

The latter conducted an examination of the extant literature pertaining to quality of life, with a specific emphasis on certain dimensions of it as opposed to considering the entirety of the concept. For example, the works [15], [16], [17] mainly focused on the use of sentimental analysis for SWB studies; the work [15] provided evidence on the importance of social media analytics identified major factors for the measurement of SWB, and presented the effects of the usage of social media on the SWB of the individuals in question. Meanwhile, [16] explored the use of big data to study SWB and concluded that emotional well-being can be measured using the tools of big data on social media data, while life satisfaction assessments cannot be valid. In addition, [17] studied the relationship between multiple aspects of hope and SWB more specifically. There is a need to analyze the different AI techniques used to date in well-being studies and thus establish an extensive state of the art to show the advantages and limitations of these approaches.

In this work, we conduct a systematic literature review on AI's potential to examine and enhance QoL studies, bridging a critical gap in interdisciplinary research. Through a comprehensive examination of research reviews spanning 2008 to 2022, we not only chart the evolution of AI applications in QoL assessments but also categorize these contributions based on their objectives, methodologies, and the QoL dimensions they target. We spotlight the integration of ML, DL, and NLP techniques in analyzing both subjective and objective QoL indicators, providing a novel classification framework that outlines the research objectives, data sources, and AI approaches employed across diverse studies. This review critically assesses the performance of these methodologies, their effectiveness, and methodological challenges, offering a comprehensive overview that highlights both the advantages and limitations of current approaches. By doing so, our work not only identifies pressing research gaps but also proposes future directions that could significantly contribute to policy-making, sustainable development, and the global endeavor to enhance well-being. The major contributions of this systematic review are:

- Interdisciplinary Bridge: We bridge the interdisciplinary gap by integrating AI with traditional QoL research domains, revealing the potential of AI in this area.
- Comprehensive Review Across QoL Subdomains: Our work represents the first to conduct a review of well-being that spans a wide array of QoL subdomains, introducing a classification that organizes these domains through the lens of AI approaches.
- Critical Analysis of AI Techniques: We provide a balanced critique of AI methods applied in QoL studies, evaluating their types, similar approaches, used datasets, effectiveness, and challenges, which illuminates the pathway for robust AI application in this domain.
- Identification of Research Gaps and Future Directions: Our work identifies unexplored areas within AI and QoL research, suggesting future research suggestions that promise to help fill these gaps.
- Practical Implications for Policy and Development: By highlighting how AI can offer nuanced insights into well-being, this work emphasizes the significance of integrating AI solutions into practical applications.

This research paper is organized as follows: In the next section, we present the state of the art of well-being studies from different perspectives and using several indexes. In the third section, we expose the research methodology that we used to select and study the research works. In the fourth section, we thoroughly analyze, study, and discuss literature findings to answer the research questions. In the fifth section, we discuss a variety of ideas concluded from our answers to the research questions, present the gaps and needs identified by the review, and discuss the potential solutions that are applicable to the well-being study. Finally, we present the key findings of our analysis as well as potential future research avenues in the conclusion.

II. RELATED WORK

In this section, we conduct a review of selected literature on the use of AI techniques in the study of QoL. The reviewed papers are summarized, and their key features, including the level of study, indicators utilized, ML algorithms applied, and other tools, are outlined.

Nowadays, AI tools, particularly ML and DL algorithms, have facilitated the analysis of QoL indicators and measures, as well as future forecasts. These tools have significantly streamlined and accelerated the processing and analysis of large amounts of heterogeneous data for researchers.

Additionally, they have proven useful in social studies, which often employ multiple methods on qualitative questionnaires to examine human behaviors and issues [18]. Specifically, supervised machine learning methods are frequently utilized to address these problems. These algorithms begin with prior knowledge to classify the studied data and are designed to predict or classify outcomes [19].

However, when traditional machine learning algorithms are insufficient for solving a problem, DL methods, which consist of neural network techniques utilizing multiple layers to represent data abstractions and build models [20], [21], are employed to improve model accuracy.

This synthesis of related work includes a comprehensive examination of numerous approaches used in the wellbeing study. We first present research works that employed sentiment analysis for studying SWB, followed by works that proposed machine learning and deep learning algorithms for measuring QoL indicators. Next, we present research works that used statistical analysis to achieve a variety of research goals. Finally, we present articles that focused on modeling QoL indicators.

A. SENTIMENTAL ANALYSIS FOR SUBJECTIVE WELLBEING STUDY

Social media data are often used to study SWB as an important indicator of QoL. These studies require natural language processing (NLP) and SA tools.

NLP allows an accurate study of the large volume of natural language text on the web. Especially since it is difficult for humans to extract knowledge from the connected world, particularly when considering time [22], SA, also known as "opinion mining," is a task within the field of NLP [23]. This task involves the identification and extraction of subjective information, such as feelings and opinions, from text through the use of text mining techniques [24], [25]. This information can be used to better understand the emotional content of the text.

SA has the potential to provide valuable insights in a range of applications, especially the analysis of human behavior and emotion through social media.

In the work [26], authors used about two billion microblogs from 1.5 million Chinese Sina Weibo users to build a method for measuring happiness using a PERMA theory-based lexicon and linguistic inquiry and word count (LIWC). The method was evaluated using the metrics of AP and Bpref and obtained scores of 0.339 and 0.543, respectively.

Furthermore, the authors of [27] used social media data from Sina Weibo to measure SWB using 1785 users' data. In this work, researchers used LIWC and the simplified Chinese version of LIWC (SCLIWC) for SA. A predictive regression model was then developed using a variety of ML algorithms, including LASSO, support vector regression (SVR), and multivariate adaptive regression splines (MARS). The latter achieved the highest Pearson correlation measure of 0.6.

While the authors of [28] studied the well-being of citizens across the United States of America (USA) by developing a regression predictive model based on georeferenced social media data using the ANEW lexicon to measure SWB, this model achieved a mean absolute error (MAE) of 0.92 and a root mean square error (RMSE) of 1.22. Its results were subsequently compared to data from the Gallup Healthways composite well-being index using Pearson correlation, and choropleth maps were created to analyze well-being at the regional and national levels.

The work [29] proposed a tool to measure SWB and specifically SWL using social media data from the myPersonality project of 101,069 Facebook users, using LIWC and the Big 5 personality for the SA task, and a two-step random forest (RF) regression model to predict SWL, as well as correlation analysis. The highest performance achieved in this work was a MAE of 0.670.

Researchers in [30] used over 35 million publications from over 20 000 Twitter users to study the SWB of citizens in Turkey by applying SentiStrength V2.2 to social media data to measure SWB and studying correlation between polarities of tweets and Gross National Happiness (GNH) results using Pearson correlation.

Another work [31] analyzed global and regional changes in SWB using a large sample of data consisting of over 2 billion tweets collected over 12 weeks from 3 geographic regions: the USA and Canada, the United Kingdom (UK), and eastern Australia. For this, LIWC and ANEW lexicons were used for SA tasks, and then PCA and Pearson correlation analysis were performed to study correlations with World Health Organization (WHO) data at the national level.

In the study [32], researchers examined the SWB of students from 44 universities in the USA, based on a sample of 79,329 tweets from Twitter. Basing on Ekman's emotions, multiple SA tools, including TF-IDF and Word n-grams, were used in addition to logistic regression and pretrained models to label data, and then Pearson correlation analysis was performed in this work as well to study the correlation of the results with data from surveys on student happiness, stress, and satisfaction. The proposed model performed the best through the utilization of TF-IDF, attaining a precision of 0.71, a recall of 0.69, and a F1 score of 0.64.

Further, the work [33] collected a two-year sample of approximately 3 billion Twitter posts to study SWL, with a focus on first-person tweets containing the keywords "me," "I," "my," or "mine." Therefore, satisfaction and dissatisfaction keywords and expressions, as well as LIWC and PERMA lexicons, were used to classify users.

The work [34] used over 10,000 publications from over 1000 Twitter users to measure SWB by employing a Naive Bayes classifier (NB) and Latent Dirichlet Allocation (LDA) to build a SA system. Additionally, the latitude and longitude were used to project the results on maps monitored over time, and the level of study is, as a result, expanded from individual to country-level.

In the research [35], Facebook data from the myPersonality project was used to study SWL and affect scores in addition to a five-item scale SWL survey for 3324 participants. Thus, researchers used LIWC and LDA for the SA task, RF for building the classification model, and finally, correlation analysis.

The authors of the work [36] investigated happiness in nine US cities by performing SA on 229,733 tweets using three polarities of sentiment scores for labeling data, which were then used to build a recurrent neural network (RNN) predictive model using the Stanford CoreNLP Toolkit.

The work [37] explored the use of SA to measure happiness, particularly with Arabic data, and developed a system capable of measuring happiness levels in multiple Saudi Arabian cities based on 2000 Arabic geo-tagged publications collected through the Twitter API. For the preprocessing task, Khoja stemmer and Mo3jam were used, as well as NRC Emotion, Sentiment Lexicons, and 3 annotators for labeling data, followed by Support Vector Machines (SVM) and NB for building the predictive model.

While the research work [38] examined happiness through the lens of social media data, specifically by analyzing locations where people were discussing specific trending topics on Twitter, a sample of 326,232 geotagged tweets across six trending topics was collected and labeled with polarity scores through the use of the SentimentIntensityAnalyzer from the Natural Language Toolkit and Hadoop for processing.

The research [39] used Twitter data from the United Arab Emirates in both Arabic and English to examine SWB. Two lexicon-based SA tools, the Hedonometer and Valence Shift Word Graphs, were utilized to analyze the data.

While the study [40] examined the relationship between fluctuations in gold prices and happiness. To assess happiness, the Hedonometer Happiness Index was employed, and Pearson correlation was used to examine the correlation between the two variables.

The work [41] investigated health and well-being indicators in the vicinity of public facilities in New York City. To this end, the researchers labeled Twitter data using the AFINN and LIWC dictionaries and examined Pearson correlations between the tweets and indicators of well-being extracted from public health and census data sources.

The authors of [42] examined national-level happiness in United Arab Emirates by utilizing Twitter data and a SA approach. The proposed model, which was based on Latent Dirichlet Allocation (LDA), demonstrated a high level of performance, with a score of 91.30% for accuracy and 0.9129 for the F-measure. Furthermore, the study also incorporated time and location as variables in order to assess changes in subjective well-being within these dimensions.

B. MACHINE LEARNING MODELS FOR QOL PREDICTION

Traditionally, QoL prediction has depended on subjective self-report measures or expert evaluations, which are susceptible to biases and limits. However, the emergence of big data and developments in ML approaches have provided new opportunities for effectively predicting and analyzing QoL outcomes.

The utilization of survey data has been a prevalent method in numerous studies aimed at investigating various indicators of well-being, including HRQoL.

These studies have relied on the collection of information through questionnaires and other forms of surveys to gather insights and better understand the well-being of individuals.

The authors in [43] conducted a survey of 669 patients and employed NLP techniques to analyze and categorize physicians' notes in order to predict HRQoL using SVM.

The research [44] used the data of 716 participants from the Korean National Health and Nutrition Examination Survey (NHANES) to build a predictive classification model of HRQoL in elderly individuals with chronic diseases by implementing multiple ML algorithms, including stepwise logistic regression, decision trees (DT), RF, and SVM. Among these algorithms, stepwise logistic regression demonstrated the highest performance, with an accuracy of 93% and an F-score of 0.49.

In the study [45], researchers used the outcomes of a socioeconomic questionnaire conducted on 200 people living in Istanbul to build a QoL index. This index's data were subsequently used to develop an ensemble ML model of various algorithms, such as SVM and MLP, and other algorithms, including logistic model tree (LMT), simple logistic regression, and the j48 algorithm, to build ensemble ML models based on stacking and voting.

In the work of [46], researchers based their research on the Oxford Happiness questionnaire and the demographic questionnaire to conduct a regression analysis to investigate the relationship between academic achievement and happiness among 180 university students.

The research work [47] identified the main factors influencing graduate students' SWB, on which the authors based their survey and studied overall happiness using two algorithms: multiple linear regression (MLR) and artificial neural networks (ANN).

The authors in [48] proposed a structural model with latent variables to predict SWL for adults of advancing age based on data from 1016 participants over the age of 55 who were enrolled in the University of Valencia's life-long learning program.

While the research of [49] developed an ensemble binary classifier using ANN, SVM, K-Nearest Neighbors (KNN), and DT to predict depression based on QoL scales using data from the NHANES in the USA.

International and regional datasets were also used in QoL studies. These datasets offered comprehensive insights on QoL based on a range of indicators across diverse regions and nations.

International indicators such as GDP, WHI, BLI, and HDI were extensively used to view QoL globally and compare between countries. These indicators provided valuable information on economic, health, and social factors affecting QoL.

The research of [50] proposed a SVM classification model to predict country-level happiness based on various QoL indicators, including the HDI and multiple human development indicators. This model is based on data collected from 187 countries by the United Nations (UN) Development Project. Based on the Myers-Briggs personality type theory, the study [51] sought to investigate multiple country-level indexes such as the HDI, GDP per capita, and democracy index. To this end, four classification models were proposed using NB, MLP, SVM, and RF.

In the research undertaken by [52], multiple supervised algorithms, including DT, RF, NN, SVR, and other linear models, were used to establish an ensemble model with the aim of predicting QoL at the country level. The study employed data sourced from the OECD's BLI from 38 countries over a four-year period. While the study [53] focused on economic well-being at the country level by predicting macroeconomic indicators using time series models and other machine learning algorithms, including LASSO, Ridge, Gradient Boosting Machines, and Random Forests, this work was based on data from the EQS Group, OECD, European Central Bank, and Bloomberg.

Additionally, authors in [54], employed different machine learning models, including ANN, SVR, and FARIMA, to forecast unemployment as an indicator of economic well-being in 22 countries utilizing the Eurostat database. Among these models, FARIMA had the best performance, with an average MAE of 0.130. The authors of [8] used multiple ML regression algorithms, including LASSO, MLR, SVR, RF, XGBoost, AdaBoost, and MLP, in order to make predictions about the WHI score of 137 countries using data from the UN Sustainable Development Solutions Network as the basis for the analysis.

ML algorithms were utilized in the analysis of health-related quality of life data, which encompasses clinical data, electronic medical records (EMRs), and data from sensors. The objective of this analysis was to assess and examine the well-being of individuals in terms of their health.

The use of AI allowed for the processing of heterogeneous datatypes and the identification of patterns and trends in HRQoL, providing valuable insights into health and well-being.

Using triangular fuzzy sets, [55] built an assessment tool for the HRQoL of people after stroke based on Bobath scale and Barthel index and other gait indicators. Further, in [56], the researchers used SVM to predict HRQoL based on an average of 114 hours of data collected from each of the 171 participants using smartphone accelerometers and gyroscope sensors.

The work [57] used multi-tasking learning and ANOVA to predict and study HRQoL based on wearable sensor data collected from 255 college students. The work of [58] used LSTM and NLP to predict life expectancy, a key indicator of HRQoL, based on 33,509 EMRs collected from seven health care facilities in Nijmegen, Netherlands.

In the study [59], clinical data from 129 alkaptonuria (AKU) patients was initially analyzed using the Pearson correlation coefficient for statistical analysis to determine the linear correlation between 11 HRQoL scores. The importance of 110 ApreciseKUre biomarkers' features was then measured using the XGBoost and then multiple ML models were developed which are KNN, linear regression, and ANN.

C. DEEP LEARNING MODELS FOR QOL

Deep learning has been used to predict quality of life, which resulted in accurate predictions and allowed to identify complex patterns and relationships in multiple types of data including questionnaires, HRQoL data, and international dataset.

The work [60] developed a fuzzy neural network neural network analysis model to assess the impact of the internet on QoL in Taiwan, using data collected from 369 participants through the WHOQoL-100 questionnaire. And in [61], the authors proposed a model to predict an individual's next-day HRQoL using ML and DL models, specifically, SVM, logistic regression, and a long-short-term memory neural network model (LSTM), utilizing data collected from the previous seven days.

In the work [62], convolutional neural networks (CNN) and analysis of variance (ANOVA) were used to investigate the relationship between radiation therapy and post-therapy functioning of the gastro-urinary system, which were measured through a HRQoL questionnaire.

Using international datasets, the research work of [63] built a multi-layer perceptron (MLP) model to investigate the relationship between multiple international QoL indicators and two air quality measures, aiming to identify the determinant well-being indicators using Eurostat data from 34 countries.

Furthermore, [64] used multiple ML algorithms, including stepwise linear regression, linear regression, DT, MLP, RF, and Gaussian Process, to analyze and identify the most influential factors on GDP growth in Bangladesh. The study used data from the World Development Indicator collected from the World Bank data repository.

D. STATISTICAL TOOLS FOR QOL ANALYSIS

Understanding and interpreting data related to QoL assessments requires statistical analysis. The analysis of quality of life comprises the investigation and evaluation of multidimensional constructs that span numerous facets of an individual's wellbeing.

In order to inform policy decisions, healthcare interventions, and social science research, it is crucial to accurately evaluate and interpret QoL data. This chapter reviews articles that primarily employ these methods to develop analytical models. Additionally, statistical tools were used in conjunction with machine ML and DL techniques to examine and explain the outcomes of predictions.

The research work [65] used a multi-dimensional approach and proposed an ordinary least squares regression (OLS) analysis model to investigate the most influential indicators on QoL using data from a proposed questionnaire.

In the research work of [66], authors employed Complex Survey Analysis of Structural Equation Models (SEM) as a statistical analysis tool to examine survey data from over 40,000 participants from 21 countries. The purpose of this analysis was to study the correlations between the composite score of multidimensional psychological well-being (MPWB) and its individual dimensions in order to provide insights for policymakers.

Furthermore, in [67], researchers used SEM for building an analysis model to study professional HRQoL and SWL of 210 Spanish nurses using data from the Healthcare Public System of the Balearic Islands.

In the study conducted by [68], a regression analysis model was proposed to examine the predictors of individuals' SWL using a large sample of 14,039 individuals aged 15 and above. And in [69], researchers conducted a regression analysis using linear mixed-effects models to investigate the impact of social emotion valuation on individual well-being and studied its correlation with national happiness scores based on a worldwide questionnaire of cognitive, emotional, and clinical indicators of SWB and data from the World Happiness Report.

Researchers in the work [70] employed a hierarchical linear regression model in their regression analysis to examine the relationships between an individual's SWB and the urbanity-level of its location based on data obtained from the National "FinSote" Survey of Finland, which included a sample of 26,422 participants.

In the work of [71], an integrated fuzzy-connective-based aggregation network with a real-valued genetic algorithm was proposed to study QoL indicators for 51 regions in the USA. While [72] conducted a regression analysis to determine the factors that have the greatest impact on the BLI and to examine the correlations between the Gender Gap Index, GDP, and BLI across 34 OECD member countries.

And in the research work [73], the authors proposed an analysis model based on multiple partial ordering tools to study and examine calculation issues related to the WHI, focusing on seven sub-indicators using data from 156 countries. In the research of [74], a regression analysis model was proposed in order to examine the rankings of 150 countries on four different indicators: the Henley Passport Index, the WHI, the Corruption Perception Index, and the Global Peace Index.

Additionally, the work of [75] proposed a regression analysis to study GDP growth using data from multiple economic indicators related to GDP in Japan. In the article [76], the authors used correlation analysis to investigate the relationships between the latitude of countries and two QoL indicators: the HDI and the WHI, additionally, the analysis was based on data from 141 countries. In the research [77], the authors used multidimensional statistical analysis to study and assess QoL for European cities based on the ISO 37120 standard.

E. SEMANTIC WEB AND OTHER APPROACHES TO MODEL QOL

In the context of investigating multidimensional QoL indicators, measurement models and semantic approaches were employed. These mainly models belong to the "semantic web" (SW), a collection of tools designed to facilitate cooperation between humans and machines [78].

Ontologies, which play a crucial role in annotating semantics and providing a common, understandable foundation for resources on the web, are key elements of this new version of the web [79]. Several studies have applied these technologies to the modeling of QoL indicators and concepts.

The research [80] proposed a model for QoL at the national level using the Multiple Indicators-Multiple Causes (MIMIC) approach. While authors in [81] built an ontology using the METHONTOLOGY approach for construction and ROMEO for validation to develop a model for national quality of life based on documents from two global indicator systems: the Global Reporting Initiative (GRI) and the OECD. Furthermore, [82] focused on QoL at the city level and proposed an ontology to integrate data from 311 sources.

In the context of smart cities and to assist with the formulation of evidence-based policies, [83] proposed "evidence-based policy making ontology" and compared it with ISO 37120 global city indicators. In another work [84], the author conducted research on the representation of city indicators and their corresponding data on the SW, specifically using Primary Student Teacher Ratio as example, resulting in the proposal of the Global City Indicator Ontology, which was validated the City of Toronto's ISO 37120 metrics reported in the year 2013. The work of [85] proposed a framework for constructing ontology models for the evaluation and similarity analysis of quality of life. The authors in [86] focused on ISO 37120 educational indicators and evaluated the data openness of city indicators in Melbourne and Toronto, both of which were implementing this index. They concluded that there is a need for semantic web tools.

The work of [87] proposed a framework for building a monitoring system in the public transportation sector at the city level. The framework is built upon an ontological depiction of the various components of the system and their connection to the Transmodel data model, which is a European benchmark for public transportation information systems.

The research work [88] focused on the concept of well-being and its various interpretations, examined how these different interpretations are often conflated, and discussed the importance of considering the specific context and cultural values when studying well-being. Additionally, they proposed the use of ontologies as a way to more accurately understand and measure well-being.

The work of [89] proposed ontologies able to represent ISO 37120 city indicator definitions. While the work of [90] proposed an ontology to help in the study of transportation, which is a major challenge and represents an important component of QoL at the city level.

In their work [91], researchers used indicators of well-being from the Better Life Index to propose a fuzzy model to assist in the formulation of strategies for a better development of life. The authors of [92] proposed an ontology for neighborhood-level sustainability that covers three domains: social, environmental, and economic. This ontology is designed to help organize and analyze data from multiple sources.

Additionally, [93] proposed an ontology for publishing and analyzing data related to education performance at the city level, using seven ISO 37120 education indicators. Other computer science tools, such as the Internet of Things (IoT), cloud storage, and blockchain, have also been used in the context of studying multidimensional QoL indicators. For instance, [94] proposed a data management system that uses these technologies to store HRQoL data in order to facilitate the measurement of quality of life.

III. RESEARCH METHODOLOGY

Literature review (LR) is a form of academic writing that summarizes and evaluates multiple research works related to a specific subject [95]. It can be characterized as an overview of a specific subject field based on well-defined research questions [96]. Systematic literature review is a type of LR with distinct objectives and characteristics. SLR follows a well-structured methodological approach [97]. In this work, we followed the process proposed by Kitchenham and Charters [98]. The purpose of this SLR is to identify, select, and review research works conducted on quality of life using data science (DS) and artificial intelligence (AI) techniques.

A. RESEARCH QUESTIONS

The primary focus of this review is to conduct a thorough assessment of the current literature related to the use of artificial intelligence to investigate quality of life indicators, aiming to pinpoint prevalent techniques, datasets, and methodologies.

Our research questions were formulated through a process that began with an initial review that helped identify key gaps and inconsistencies, setting the stage for collaborative question development aligned with our study's objectives. This development was guided by essential data science steps commonly adopted in scholarly research, which involved: defining research goals, selecting suitable data, using rigorous data preprocessing techniques, employing machine learning algorithms and statistical tools, or applying modeling approaches in other works.

These methodical steps helped to refine our questions, ensuring they were precise and positioned to yield significant insights. This resulted in the formulation of these five research questions:

- RQ1: What are the objectives of QoL research?
- RQ2: What datatypes, sources, and features were used in this context?
- RQ3: Which preprocessing techniques were employed during the data preparation phase?
- RQ4: Which machine learning algorithms and statistical tools were utilized?
- RQ5: What approaches were used to model QoL indicators?

B. SEARCH STRATEGY

We use multiple websites and digital databases, including Scopus, Elsevier, PLOS, ACM, SpringerLink, the IEEE Xplore digital library, the DBLP database, and PubMed. We defined the appropriate keywords in order to extract the maximum number of articles related to the topics of our research. These keywords were mainly quality of life synonyms, abbreviations, and words related to it, as well as some of its important sub-components such as "happiness," "well-being," "subjective well-being," "quality of life," "satisfaction with life," "standards of living," "life satisfaction," "happiness index," "health-related quality of life," "QoL," and "HRQoL." Subsequently, we constructed search strings based on the keywords. Afterwards, we interrogated the aforementioned digital databases, extracting articles that pertained to our area of investigation.

C. STUDY SELECTION AND QUALITY ASSESSMENT

With the goal of selecting the most relevant articles related to our topics, we specified the inclusion and exclusion criteria that are described in Table 1 and the steps taken during the pre-screening process are depicted in Figure 1.

Upon excluding duplicates and ensuring compliance with the exclusion criteria, the selected articles underwent a quality assessment process. This involved, firstly, the evaluation of the presence of related keywords to the specified topic in

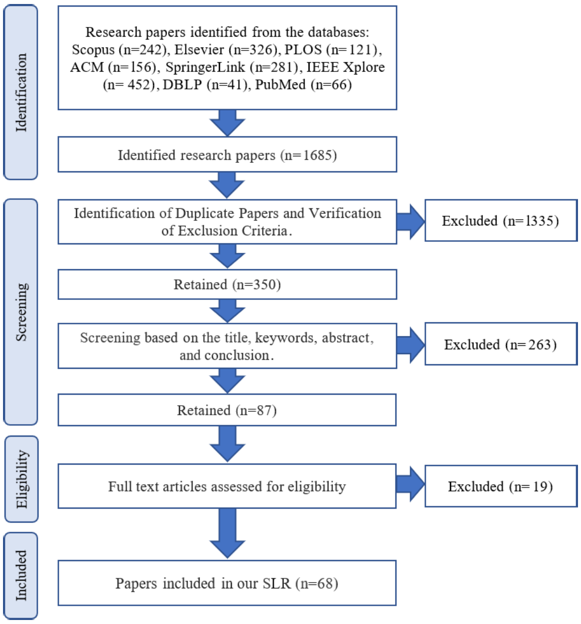

FIGURE 1. PRISMA flow diagram.

TABLE 1. Inclusion and exclusion criteria used to select relevant articles to this review.

Inclusion criteria	Exclusion criteria
The articles are written in English.	Irrelevant articles to our topic and research questions.
The articles are from books, journal publications, or conferences.	The articles not dealing with the concept of well-being
The works that are indexed in Scopus, IEEE, ACM, Elsevier, Science Direct, or Springer.	The articles not in the range of 2008 to 2022.
The works related to the study of happiness and quality of life, with their various types and definitions	The articles that did not use data mining, a computer science tool, or a statistical method.
The works that proposed a data driven solution for QoL using data science, statistics, or modeling tools.	Duplicate articles.

the title and abstract of each article. Secondly, the articles were analyzed for relevance of content from the abstracts and conclusions and for addressing the six research questions. The evaluation process was a collaborative effort, with all authors contributing to the analysis of the keywords and main objectives of the research works, and conducting a full-text review to determine their relevance to the specified topics.

D. DATA ANALYSIS

The comprehensive analysis of selected studies across various publication types and venues provides a nuanced understanding of the evolving landscape of research in well-being through the lens of DS and AI as presented in Table 2. This diverse collection, ranging from book sections and conference papers to journal articles, underscores the multidisciplinary nature of well-being research, highlighting the convergence of technology, health, and social sciences.

Book sections published by Springer, such as [27] and [36], offer in-depth insights into specific technological advancements and their applications in studying well-being. These chapters contribute significantly to the academic discourse by providing comprehensive overviews, theoretical frameworks, and case studies.

Conference papers from prestigious organizations like ACM and IEEE demonstrate the dynamic and innovative nature of ongoing research in this field. Papers presented at conferences such as [28] and [46] showcase cutting-edge research, methodologies, and findings.

Journal articles published by entities like Elsevier, IEEE, and Nature Portfolio, among others, offer rigorous peer-reviewed research findings that enrich our understanding of well-being from various perspectives. Studies in journals such as [53] and [69] not only provide empirical evidence and advanced analytical techniques but also explore theoretical dimensions of well-being. These articles are instrumental in advancing knowledge, informing practice, and shaping policy decisions by presenting scientifically validated insights into the multifaceted aspects of well-being.

The inclusion of diverse publication types and publishers illustrates the broad interest and investment in researching well-being through DS and AI. This multidisciplinary approach is essential for tackling the complex, multifaceted nature of well-being, which encompasses physical, mental, and social dimensions. The selected studies span a wide range of topics, from the application of big data in public health to the development of AI-driven interventions for individual well-being. This diversity not only reflects the richness of the field but also highlights the potential for DS and AI to contribute to our understanding and improvement of well-being at various levels.

Moreover, the temporal range of publications, from 2008 to 2022, indicates the rapid evolution and growing sophistication of research methodologies and technologies in this area. The progression from early explorations of AI and big data in well-being to more recent studies employing advanced machine learning algorithms and comprehensive data analytics reflects the technological advances and increased availability of data. This evolution underscores the importance of interdisciplinary research and collaboration in harnessing the full potential of DS and AI for well-being.

While figure 2 depicts the distribution of publications based on their publication dates. Approximately half the articles are clustered between 2019 and 2022, underlining the importance of investigating quality of life indicators in that particular period.

In next chapters we will dive into the analysis of selected studies to highlight the significant contributions of DS and AI to well-being research, and the importance of interdisciplinary approaches and continuous innovation. This body of work not only contributes to academic knowledge but also has the potential to inform policy, guide practice, and ultimately enhance the well-being of individuals and communities worldwide.

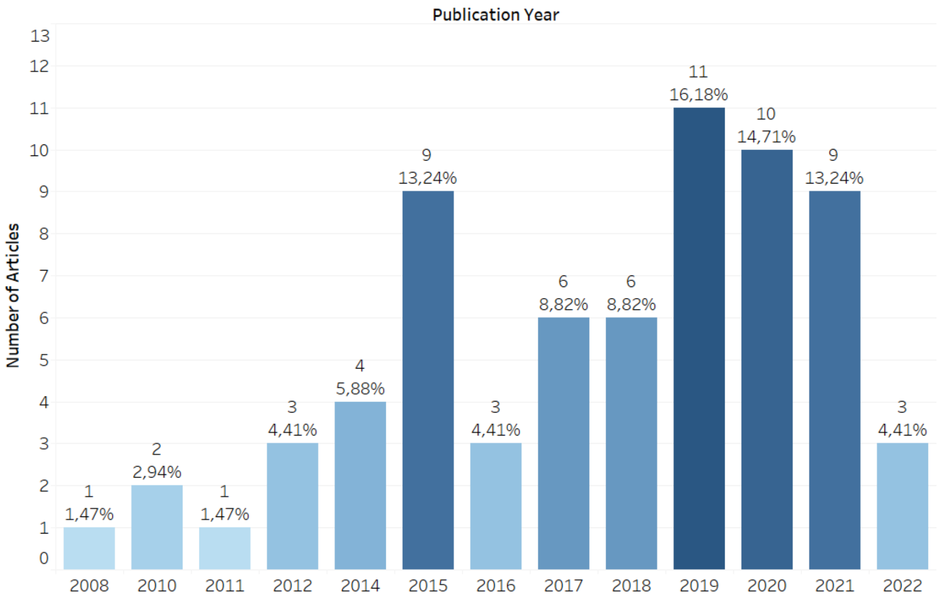

FIGURE 2. **Distribution of articles by year of publication.**

IV. REVIEW ANALYSIS RESULTS

Artificial intelligence has the potential to make significant contributions to the study of quality of life by analyzing large datasets, developing personalized recommendations and interventions, automating the collection of data, and enhancing our understanding of this concept. QoL is a multi-faceted concept that is studied using multiple fields, including psychology, sociology, economics, public health, and urban planning.

A. QOL RESEARCH OBJECTIVES (RQ1)

In this section, we outline the objectives of the selected research as well as the types of AI approaches that have been employed in accordance with the research goals. These approaches encompass either measuring a well-being index, analyzing its outcomes, or proposing a model of an indicator or a set of dimensions for the quality of life.

It is important to have a clear definition of the concept of ''quality of life'' in order to effectively study and measure it. In this literature review, we identified two criteria for projecting quality of life indicators: the level of study and the QoL dimensions which represent the objective of studies in this context.

For the first criteria, we discovered that ''quality of life'' is not always studied as a concept covering all fields related to the measurement of well-being. Rather, in multiple works, the authors adopted a clearly stated dimension for their studies. These dimensions can include:

- First, health-related quality of life is related to the assessment of individuals' health status.
- Second, subjective well-being: the study and measurement of ''happiness'' at all levels is an important sentiment analysis task.
- Third, quality of urban life: an indicator used to study cities' well-being and urban services, especially using the ISO 37120 model.
- Fourth, economic well-being: this dimension is related to studying the economic indicators of countries such as GDP and unemployment rates.
- Fifth, in this context, we refer to ''general well-being'' when the indicators include a range of aspects belonging to multiple fields of study and which can be both objective and subjective. Further, this is related to the measurement and study of well-being at the country level, which should consider all dimensions that can help in the assessment of citizens' standards of living, whether they are subjective or objective indicators. The most commonly used general well-being indicators at country level are the WHI, the BLI and the Human Development Index.

While considering the second criteria, QoL indicators are studied at multiple levels:

(1) Individual-level: This is the most studied level in the literature and focuses on the state of well-being of individual people, particularly SWB and HRQoL.
(2) City-level: Frequently referred to as the ''quality of urban life,'' this is often investigated in order to model quality of life indicators at this level of study.
(3) Country-level: Research at this level is highly diverse and involves the use of national and international well-being indices to study, measure, and model quality of life indicators. This is the second-most researched level, indicating the importance of evaluating well-being at the country level.

Figure 3 displays the results of projections on the hierarchy of the QoL concept and its 5 research objectives,

TABLE 2. Inclusion and exclusion criteria used to select relevant articles to this review.

Type	Publisher	Publication Title	Selected studies
Book section	Springer	Active Media Technology	[27]
		Advanced Informatics for Computing Research	[34]
		Artificial Intelligence and Soft Computing	[55]
		Artificial Intelligence Applications and Innovations	[63]
		Big Data – BigData 2018	[36]
		Computational Collective Intelligence. Technologies and Applications	[60]
		Social Computing, Behavioral-Cultural Modeling, and Prediction	[29]
		Social Informatics	[32]
Conference paper	ACM	Proceedings of the 30th Annual ACM Symposium on Applied Computing	[28]
		Proceedings of the Seventeenth Australasian Document Computing Symposium on - ADCS '12	[81]
	IEEE	2014 11th Web Information System and Application Conference	[26]
		2015 IEEE First International Smart Cities Conference (ISC2)	[83] [86]
		2015 International Conference on Orange Technologies (ICOT)	[72]
		2018 1st International Conference on Computer Applications & Information Security (ICCAIS)	[37]
		2018 IEEE 3rd International Workshops on Foundations and Applications of Self* Systems (FAS*W)	[94]
		2018 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)	[51]
		2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)	[38]
		2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)	[57]
		2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)	[61]
		2019 IEEE 11th International Conference on Engineering Education (ICEED)	[46]
		2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)	[64]
		2021 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT)	[8]
	KSI Research	The 27th International Conference on Software Engineering and Knowledge Engineering	[50]
	Scopus	AAAI Workshop - Technical Report	[82]
		CEUR Workshop Proceedings	[90]
		Proceedings of the International Workshop on Semantic Big Data, SBD 2020 - In conjunction with the 2020 ACM SIGMOD/PODS Conference	[92]
Journal article	BioMed Central	Orphanet Journal of Rare Diseases	[59]
	Elsevier	European Journal of Operational Research	[53]
		Finance Research Letters	[40]
		Procedia - Social and Behavioral Sciences	[65]
		Technological Forecasting and Social Change	[30]
		The Journal of Socio-Economics	[80]
	IEEE	IEEE Journal of Biomedical and Health Informatics	[31], [56]
	Inderscience Publishers	International Journal of Applied Decision Sciences	[47]
	MDPI	Applied Sciences	[52]
		International Journal of Environmental Research and Public Health	[67]
		Sustainability	[76]
	Nature Portfolio	Scientific Reports	[69]
	PLOS	PLOS ONE	[33], [35], [77]
	PubMed	AMIA Annual Symposium	[43]
		Computers in Biology and Medicine	[62]
		Health and Quality of Life Outcomes	[66]
	SAGE Journals	Progress in Human Geography	[88]
	Science Direct	Computers, Environment and Urban Systems	[84]
		Expert Systems with Applications	[93]
		Journal of Web Semantics	[89]
	Scopus	International Journal for Quality Research	[85]
		Transportation Research Part C: Emerging Technologies	[87]
	Springer	Applied Research in Quality of Life	[41]
		BMC Medical Informatics and Decision Making	[58]
		Brain Informatics	[49]
		Computational Economics	[54], [75]
		Current Psychology	[68]
		International Journal of Fuzzy Systems	[91]
		Journal of Ambient Intelligence and Humanized Computing	[45]
		Journal of Big Data	[39]
		Journal of Happiness Studies	[48]
		Journal of International Migration and Integration	[74]
		Neural Computing and Applications	[71]
		Social Indicators Research	[70]
		Social Network Analysis and Mining	[42]
		Sustainability Science	[73]
	KOSMI	Healthcare Informatics Research	[44]

which are linked to the three levels of study. The figure also shows examples of the used indicators associated with each objective.

These QoL dimensions are the research goals of the studies analyzed in this systematic review. Figure 4 shows how different approaches are used to fulfill these research goals. We note that the majority of studies are devoted to measuring and predicting indicators of well-being at various levels. Individual level is the most targeted with 21 articles, mostly concerned with studying HRQoL and SWB indicators.

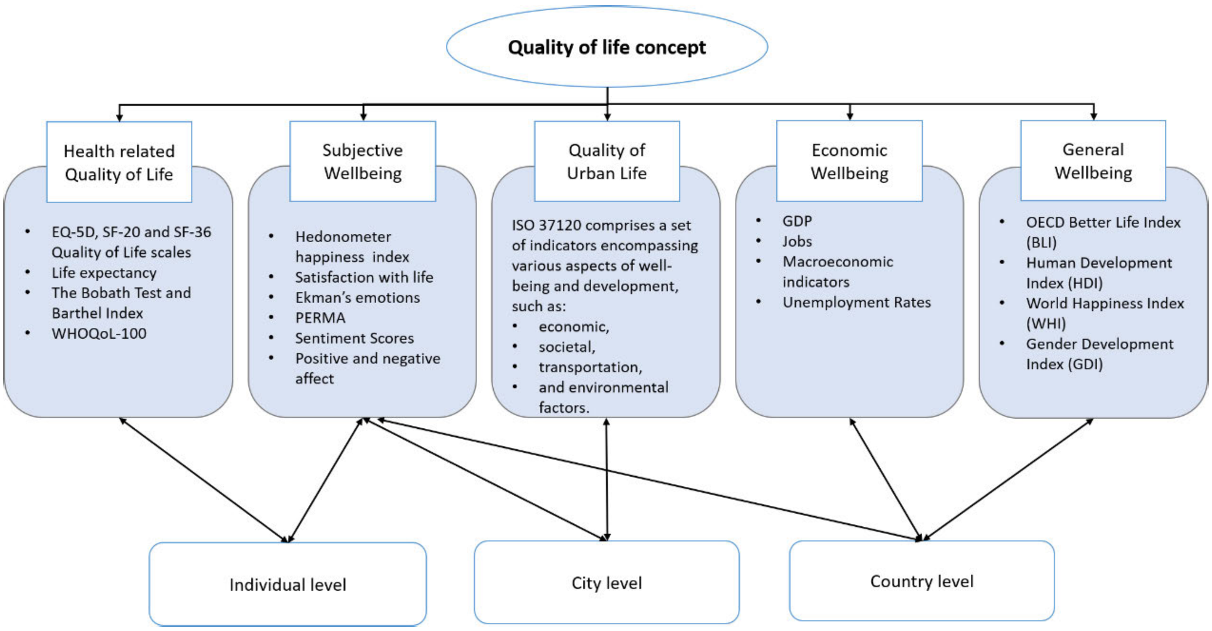

FIGURE 3. QoL study objectives according to their levels of study.

Furthermore, the SWB study, which is primarily concerned with individuals, can be related to level expansions at other levels as well. Next comes the country level with 20 articles. We observe that general well-being at this level is the most commonly used for QoL analysis. This indicates the importance of using AI techniques for capturing macro-level trends and patterns at the international level and revealing important factors affecting wellbeing. While QoL modeling approaches were employed to model QoL at different levels, with a focus on city-level indicators, in conclusion, 37 research works proposed predictive models to measure and study QoL, and 16 articles were dedicated to the analysis of QoL indicators. Additionally, 15 papers were dedicated to modeling QoL concepts.

B. DATA, METADATA, AND FEATURES (RQ2)

1) DATA USE CASES

With the aim of applying the aforementioned approaches used to study QoL, researchers employed a variety of data types. First, we project these data types on goals of study as the figure below illustrates.

Figure 5 shows the types of data used to study each QoL dimension. Questionnaires are generally the most common data type in the literature. This data type is used to achieve a diversity of research goals, mainly related to subjective well-being and HRQoL indicators. Next, social media data are the second most commonly used data type in the selected literature; a social media source is a collection of posts published by users of microblogging sites.

These posts generally contain details on how people feel about their lives at the moment of publication, which makes this type of data very rich with information about individuals' satisfaction with life and their subjective well-being. International datasets were also used, and they are often related to a QoL index, especially at the country level since the datasets are related to national and international indicators of well-being. As we observe in Figure 5, this data type is used in the study of general well-being and in some works examining economic well-being. Clinical data, EMRs, and sensors are a set of data types used in research to measure HRQoL using machine learning algorithms. It is worth mentioning that the building of QoL models primarily relied on non-data-driven methodologies.

2) DATATYPES, SOURCES, AND SIZES

In the realm of studying the quality of life using diverse data types, it's essential to recognize that there is no one-size-fits-all approach when it comes to acquiring the necessary data. This is primarily because the subject matter is multifaceted and dynamic, encompassing a wide range of factors that contribute to an individual's or a community's well-being. As a result, researchers and analysts often find themselves navigating a complex landscape of data collection methods and sources. Here, we delve into the intricacies of data acquisition for QoL studies:

Unlike some well-defined fields of study that have established standard datasets, quality of life research lacks universally accepted datasets. Quality of life encompasses numerous dimensions, including economic, social, environmental, and health factors. These dimensions often vary from one study to another, making it challenging to create standardized datasets that suit all research objectives.

Some valuable datasets related to quality of life are provided by international organizations such as the United

FIGURE 4. QoL study objectives, used approach and level of study.

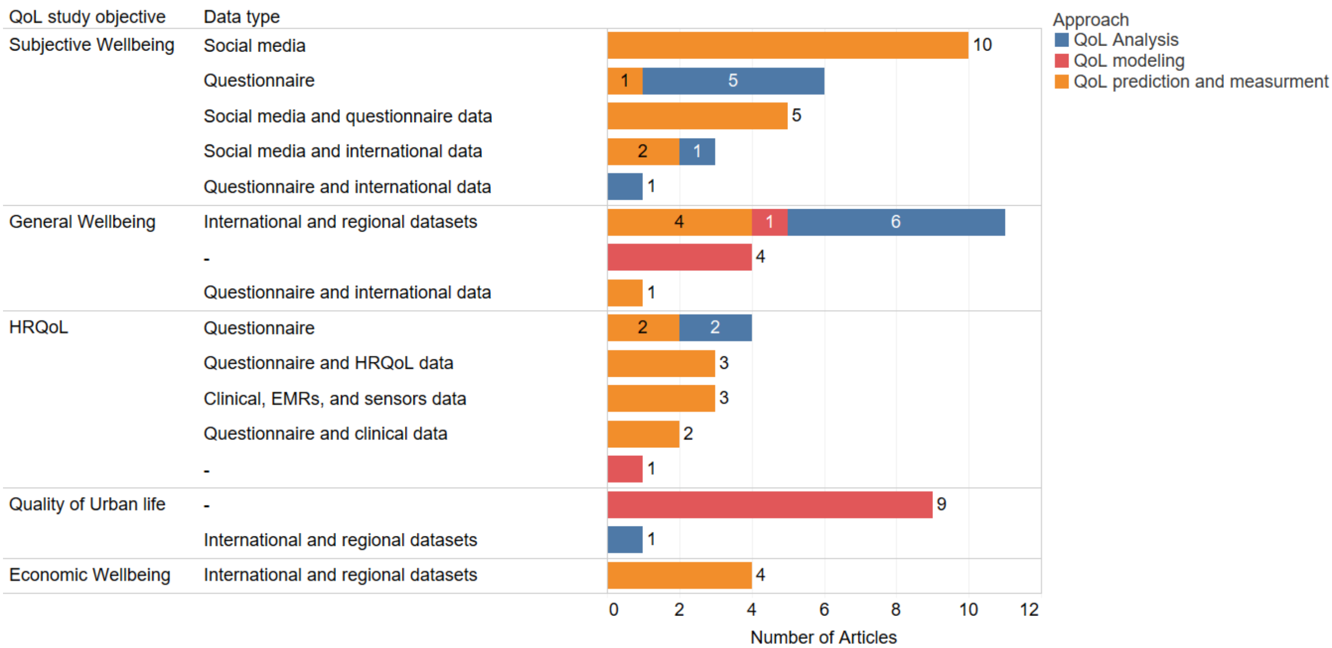

FIGURE 5. QoL study objectives and their according used data types.

Nations, World Bank, OECD, and others which are used in these works [8], [50], [51], [52], [53], [54], [63], [64], [69], [72], [73], [74], [75], [76], [81]. These organizations compile and publish data on a wide range of indicators, including income, education, healthcare, and environmental quality. Researchers often rely on these datasets as they are comprehensive, reliable, and internationally recognized. Table 3 shows that five indicators' datasets were used in more than one research work.

To address the absence of standard datasets, many researchers resort to web scraping. This involves programmatically extracting data from websites, databases, and online sources. Additionally, depending on the specific research objectives, scholars may conduct custom surveys and studies to gather data on quality of life. Thus, in the research works [35], [43], [45], [62], [65], [69], [91], as depicted in Table 3, researchers proposed custom surveys to capture unique aspects of well-being not covered by existing datasets.

Custom data collection methods allow researchers to tailor their research to the local context or specific population they are studying, for instance, in [32], [47], [48], and [61] researchers targeted college students and designed adequate surveys to assess specific indicators related to the studied sample. i Collecting data from social media on individuals' perceptions is vital for its real-time, unfiltered insights. With its vast user base, social media provides a diverse pool of perspectives, enabling comprehensive analysis. Natural language processing further streamlines data extraction, facilitating both immediate and longitudinal studies. This approach proves cost-effective compared to traditional methods, reducing recruitment efforts and expenses.

We conclude that in order to achieve QoL research goals on multiple dimensions, it is related to utilizing a variety of data types in multiple works. Upon studying the literature, we found that these types of data can be categorized into four types: 1) international datasets; 2) questionnaires; 3) social media; and 4) clinical, EMR, and sensor data. Table 3 shows the datasets for each data type. Moreover, we observe that Twitter is the most used social media data source, while WHI, BLI, and HDI are the most used international datasets. Additionally, we note that, for both HRQoL and questionnaire data, there are almost no commonly used datasets.

The diversity in data types underscores the multi-faceted nature of QoL assessment. The size of the studied samples varies significantly across diverse locations and time periods among the collected works.

In Table 4, and for a better interpretation of findings, we included specific locations and time periods extracted from the studied literature in addition to numbers of rows and participants to highlight the distinctions among data types and emphasize the imprudence of studying quality of life dimensions in isolation from various vantage points. In this table, we present the data-based contributions projected on the four data types in addition to four standards for measuring the size of the data, which are the number of publications, people, or countries depending on the magnitude of the study and the associated type of data, in addition to the time periods included in the studies, which is important information, especially when working on timeseries data.

We note that HRQoL and questionnaire data, which are the same in some cases, are often quantified by the number of participants, while international dataset size is defined in terms of geographical location and time period. The social media-collected datasets can be compared by considering all four projections. However, it is essential to consider the potential limitations and biases inherent in each data source when drawing conclusions or making policy recommendations.

3) FEATURES AND FEATURE TYPES

In this section, we embark on a detailed exploration of the various dimensions and categories that encapsulate the vast domain of QoL. Recognizing the complexity and the multifaceted nature of well-being, we focus on dissecting the different feature types that researchers have scrutinized to understand the nuances of human well-being.

We contribute to the research by studying the features used in the examined works and categorizing them into distinct classes. We aim to build a structured lens through which the breadth and depth of well-being research can be examined. To systematically approach this complexity, we have organized the features into ten primary categories. These encompass a wide range of aspects from health and mental well-being to the impact of politics and governance on individual and societal wellness. Each category has been meticulously examined through different datatypes, including clinical data, surveys, and international datasets, to paint a holistic picture of the current research landscape, as shown in both Table 5 and Figure 6.

We note from the figure that health-related features have seen a relatively balanced research approach across three data types, that encompasses clinical data, surveys, and international studies. This reflects the multifaceted nature of health as a component of well-being.

Additionally, the figure illustrates that the most frequently studied features are SWB and other social media features, particularly through questionnaires, surveys and social media data. Economic features have also drawn attention primarily through international and regional datasets. Which highlights the usage of macro-level approaches to understanding economic well-being.

Conversely, education and environment feature types have been the subject of fewer studies, this may indicate gaps in the current research landscape. These areas might benefit from increased scholarly attention to provide a more comprehensive understanding of their roles in human well-being.

The study of wellbeing through the application of data science tools encompasses a multifaceted examination of various key features. In addition to Figure 6, Table 5 is also presented to provide a more detailed breakdown of feature categories and their subcategories.

This table lists each of the categories in the left column and provides a brief description of each category. It also includes the subcategories that fall under each category, which provides a more comprehensive understanding of the types of features studied in the works analyzed.

(1) In the domain of health and healthcare, comprehensive datasets comprising CT scans, sensors' data, and EMRs play a pivotal role, in addition to other features extracted through surveys to assess health status. These data help facilitate the early detection of diseases and enable the tailoring of treatments to individual patients, ultimately resulting in improved healthcare outcomes.

(2) Demographic information, encompassing factors such as age, gender, and location, emerges as a crucial component in understanding disparities in studied samples. This foundational data serves as a compass, guiding targeted interventions to address specific needs within diverse populations.

TABLE 3. Datasets used to study the concept of QoL according to their types by reference.

Data Type	Dataset	References
International and regional datasets	World Happiness index data	[73], [8], [74], [76], [69]
	OECD's Better life index data	[81], [72], [52], [53]
	Human development indicators	[50], [52], [76]
	Eurostat database	[63], [72], [54]
	GDP per capita	[51], [75]
	World Bank data	[64]
	Bloomberg data	[53]
	Corruption Perception Index and Global Peace Index data	[74]
	London bullion market gold prices	[40]
	WHO data	[31]
	Financial news from the website of the EQS Group	[53]
	Regional data in the United States	[71]
	ISO 73120 World Council on City Data	[77]
	Data from The Center for Disease Control's Behavioral Risk Factor Surveillance System and the New York City Department of Health and Mental Hygiene	[41]
Questionnaire data	A proposed survey	[43], [65], [35], [91], [45], [62], [69]
	Survey on College students	[32], [47], [48], [61]
	National Health and Nutrition Examination Survey (NHANES)	[44], [49]
	Survey results of Turkish Statistical Institute	[30]
	Big 5 personality questionnaire	[29]
	The demographic and Oxford Happiness Questionnaires	[46]
	PANAS and PWBS survey	[27]
	The European Social Survey	[66]
	The Gallup World Poll	[68]
	Spanish nurses' data from a Healthcare Public System	[67]
	The National FinSote Survey	[70]
	WHOQoL-100	[60]
Social Media data	Twitter dataset	[30], [31], [28], [32], [33], [34], [37], [36], [39], [38], [40], [41], [42]
	Facebook & myPersonality project dataset	[29], [35]
	Sina Weibo dataset	[26], [27]
HRQoL data	Sensors and HRQoL data of college students	[57]
	Clinical notes	[43], [58]
	Bobath scale and Barthel index data from post-stroke patients	[55]
	Smartphone accelerometer and gyroscope sensors data	[56]
	The ApreciseKUre database	[59]

Referring to Figure 6, it is worth mentioning that these characteristics were employed as supplementary details alongside information pertaining to a group of individuals.

This included data from social media, the initial section of the questionnaires, as well as essential patient information in the conducted research.

(3) Furthermore, subjective wellbeing and mental health data contribute invaluable insights into an individual's emotional state, encompassing aspects such as happiness, stress levels, and overall life satisfaction. SWB features are used in the four data types, but mostly in social media and questionnaire data.

(4) Economic indicators, such as GDP per capita and employment rates, are indispensable in understanding the broader socioeconomic context that influences overall wellbeing. These metrics reflect the financial stability of a region or country and serve as a barometer of access to education, healthcare, and essential resources. These attributes have predominantly been examined within an international framework across 10 studies. In the case of questionnaires, socioeconomic investigations focused on individual and household factors such as income and unemployment.

(5) Social and community dimensions are equally pivotal. Data on social support networks and community engagement represent the influence of social connections on an individual's sense of belonging and wellbeing. Understanding these dynamics empowers communities to develop strategies

TABLE 4. Datasets used to study the concept of QoL and their size by datatype and reference.

Data Type	QoL dimension	Ref	Data size			
			Number of rows	Participants	Location	Time period
HRQoL Data	HRQoL	[58]	33,509 EMRs			
		[55]		40 patients		
		[56]		171 participants		
		[59]		129 patients		
		[43]		669 patients		
International and regional datasets	Economic Wellbeing	[75]	147		Japan	Q4 1981 – Q2 2018
		[53]	80,813			1996-2016
		[54]			22 countries	2000-2014
		[64]			Bangladesh	1991-2018
	General Wellbeing	[72]			34 countries	2011-2015
		[50]			187 countries	
		[51]			124 countries	
		[76]			141 countries	1990-2019
		[81]			Multiple countries	
		[63]			34 countries	
		[74]			150 countries	
		[52]			38 countries	2014-2017
		[71]			51 states, USA	
		[8]			137 countries	2015-2021
		[73]			156 countries	2016
	QoUL	[77]			Multiple European cities	
	Subjective Wellbeing	[69]		7,443 participants	40 countries	2019
		[68]		14,039 individuals	Italy	2005-2017
Questionnaire	HRQoL	[49]			USA	2013–2014
		[67]		210 nurses	Spain	
		[44]		716 individuals		2008-2010
		[60]		369 participants		
		[62]		52 patients		
		[61]		142 participants		30 days
		[57]		255 students	New England, USA	2013-2017
	Subjective Wellbeing	[47]		118 students		
		[48]		1016 people		
		[65]		165 participants		
		[46]		180 students		
		[66]		41,825 participants	21 countries	
		[70]		26,422 participants		
	General Wellbeing	[45]		200 people	Istanbul, Turkey	
Social media	Subjective Wellbeing	[26]	2 billion microblogs	1.5 million		
		[37]	2000 tweets			January-June 2018
		[42]	+0.3 million tweets daily.		United Arab Emirates	September 2017
		[28]	500,000 tweets		USA regions	2012
		[38]	326,232 tweets			February-March 2018
		[36]	229,733 tweets		9 cities	3 days
		[39]	+17 milions			
		[34]	+10 000	+1000 individuals		
		[41]	2,731,504 tweets		New York, USA	December 2015-June 2016
		[31]	2.73×10^9 tweets		USA, Canada, UK, and Australia.	12 weeks
		[30]	+35 million tweets	+20 000 individuals	Turkey	
		[40]	2871			September 2008-December 2019
		[29]		101,069 users' data		
		[35]		3324 participants		2009-2011
		[27]		1785 users' data		
		[33]	3 billion tweets			2012-2014
		[32]	79 329 tweets	133 students		

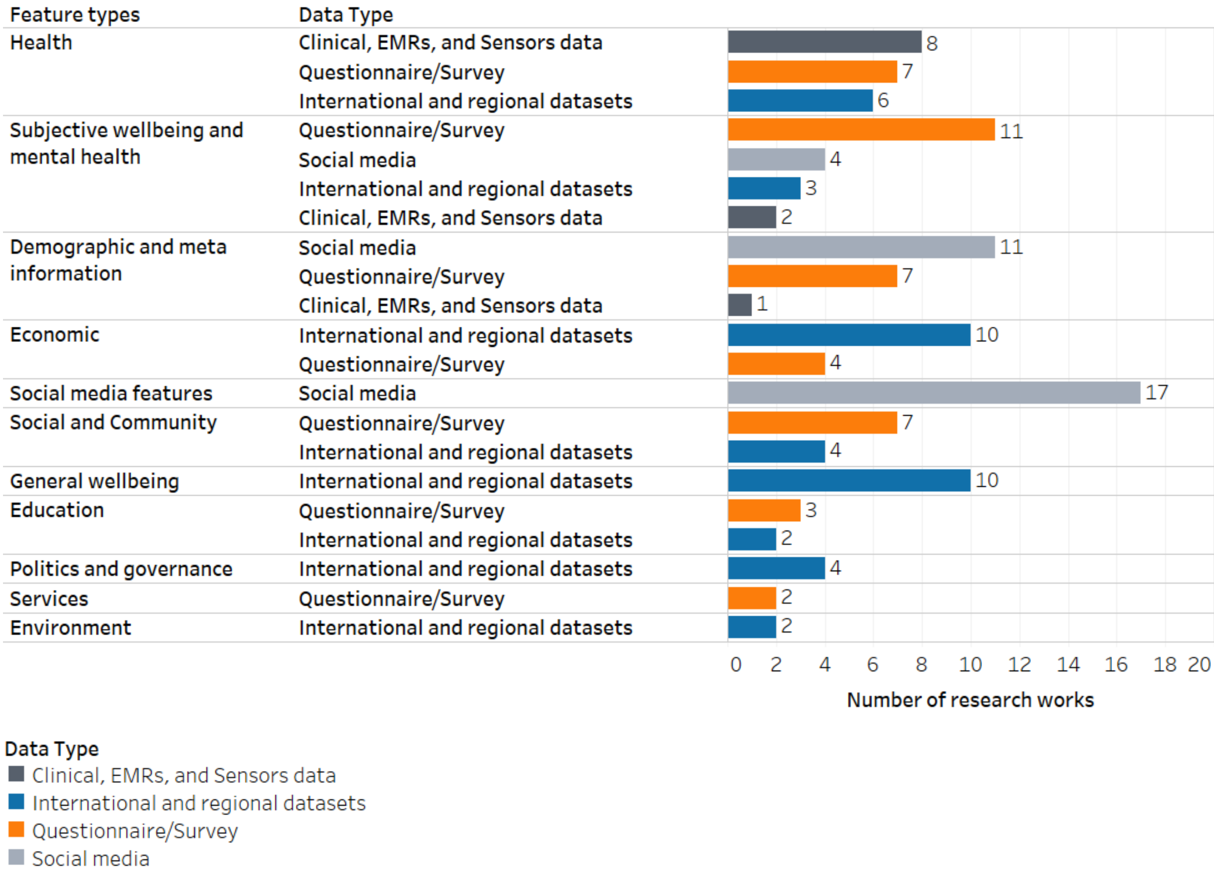

FIGURE 6. Features used in studied works according to each data type.

TABLE 5. Detailed features for each feature type.

Feature types	Features
Health	CT scans, EMR, gait and mobility, health biomarkers, healthy life expectancy, clinical notes, infant mortality, life expectancy, maternal mortality ratio, pain and discomfort, physical health issues, physiological data, public health and healthcare services, self-care, sensor data, and vitality.
Demographic and meta information	Age, gender, time, location, relationship status, and other user information
Subjective wellbeing and mental health	Subjective wellbeing, achievement, affect, happiness scores, anxiety/depression, attitudes, autonomy items, environmental mastery, personal growth, conscientiousness, and neuroticism. Life satisfaction, meaning, and purpose in life. Mental health, personality types, self-acceptance, stress.
Economic	Consumer Price Index, inflation, trading data, dependency ratio, economic news, employment/unemployment/jobs, financial burden, GDP per capita, GDP growth, gold prices, gross national income, housing and stock market data, income, USD exchange rate, and monetary data.
Social and Community	Extraversion and openness, social support, community, family, engagement, generosity, social and cultural issues, social indicators, relationships, and social services.
Social media features	Events, topics, tweets, and publications
General wellbeing	Better Life Index, World Happiness Index, Human Development Indicators, Gender and Human Inequality, Gross National Happiness, Standard of Living, and other aspects of wellbeing.
Education	Academic discourse, expected years of schooling, female secondary education, and youth literacy rate
Politics and governance	Corruption Perception Index, Global Peace Index data, Democracy Index, and other political dimensions.
Environment	Environmental sustainability, emissions of sulfur oxides and nitrogen oxides, and other environmental aspects

that strengthen bonds, enhance social cohesion, and promote collective wellbeing.

(6) The analysis of social media data provides a unique lens into public sentiment and interests, offers valuable insights

into the collective consciousness of a society, and gives the ability to discern trends, concerns, and areas of focus within the broader community.

(7) Through the use of general wellbeing indicators, policymakers can gain a holistic view of a nation's overall wellbeing and compare multiple countries. Examples of these indicators are included in this table.

(8) Metrics related to education allow us to understand how access to learning impacts an individual's opportunities, job prospects, and income.

(9) Data pertaining to corruption levels and the effectiveness of governance systems are foundational in shaping transparent and accountable government institutions. These elements directly influence the equitable distribution of resources and the delivery of public services.

(10) Environmental data illuminate the far-reaching impacts of sustainable practices and policies on air and water quality, public health, and overall quality of life. This information is crucial in guiding efforts towards conservation and sustainable development.

C. PREPROCESSING AND FEATURE ENGINEERING (RQ3)

1) NLP AND SA FOR LABELING AND PREPROCESSING

When dealing with objective concepts included in QoL indexes, experts from every discipline can resolve the ambiguity of the QoL indicators relevant to their field of expertise by defining objective measures. On the other hand, the challenge is quantifying and measuring subjective well-being; therefore, various research works on well-being have been dedicated to using natural language processing tools for studying SWB and extracting important information from textual data.

Social media data are the most commonly used datatype, specifically to measure and study subjective well-being. Twitter is the most used source for collecting data due to its available API. Facebook data from my personality project was used in two papers, as well as Sina Weibo, which is a Chinese microblogging website. Additionally, NLP approaches were applied to clinical notes and news articles as well. Social media platforms generate vast amounts of data in real-time, and NLP tools are used to process and analyze this data and identify sentiment, trends, and user behavior. These tools can be applied to clinical notes to extract and analyze information that is relevant to healthcare professionals. For example, to identify patient symptoms, diagnose diseases, and determine the effectiveness of treatments, or to extract relevant information from electronic health records, which is used to improve patient care and outcomes.

A variety of natural language processing tools were used in the context of studying wellbeing, as described in Table 6, for cleaning, feature selection and extraction, annotation, text analysis, and topic modeling tasks.

Feature extraction is used for extracting new features from existing data. These include techniques such as n-grams and different types of word embeddings such as count/frequency-based techniques, which include TF-IDF and word count, in addition to prediction-based techniques such as Word2Vec and GloVe. Word entropy is effective for text classification and topic modeling. Each technique has advantages and disadvantages, and the technique used depends on the task at hand and the nature of the text data. These strategies are crucial for developing NLP models that can extract insights and information from textual data.

Labeling or tagging text data is essential for training machine learning algorithms that can automatically recognize and classify text data. This improves the accuracy and efficiency of NLP models and enables applications like text classification and sentiment analysis. Mainly, this is achieved using a variety of sentiment lexicons and dictionaries that contain words or phrases associated with labels such as positive and negative sentiment.

Next, text analysis methods were widely used for studying textual data, especially LIWC. These types of NLP methods enable researchers and analysts to extract insights from large volumes of text data by identifying patterns and trends in text data that may not be immediately obvious.

Topic modeling was also used for identifying the most important topics or themes in the studied textual data in question, as it is essential for quickly and accurately identifying the most important themes in a large volume of text data.

Overall, the most used tools are LIWC for text analysis, followed by TF-IDF for feature extraction. The Hedonometer and ANEW are the most used lexicons, both of which are used in the labeling task.

Given the prevalence of data related to SWB on social media platforms, it is possible to conduct research on large samples and expand the scope of analysis beyond the individual level to include regional and city-level as proposed in [30], [32], [36], [37], [38], and [41] and country-level as proposed by [26], [28], [31], [34], and [42].

Geotagged data, which includes information on the location of social media users, enables researchers to expand their analysis to include larger geographical levels, as illustrated in Figure 7.

By incorporating location data, researchers are able to track changes in happiness in different geographical locations. Additionally, the date of publication of social media posts can serve as a temporal dimension, allowing researchers to examine changes in SWB over time.

2) STATISTICAL AND ML PREPROCESSING TOOLS

Table 7 shows the preprocessing and feature engineering methods that were used in the chosen quality-of-life studies. It shows the techniques that were used for each task and the source of each study. It is noteworthy that most of the works presented in this table achieved the best performances among the studied works and will be presented in the next sections.

For social media data, the primary focus lies on NLP tools, which were covered in the previous part, given the unstructured and often noisy nature of such data. Thus,

TABLE 6. Natural language processing methods applied in QoL studies.

Natural language processing task	Methods	Research Works
Cleaning and Preprocessing	Bayesian nets	[42]
	Chinese Word Segmentation	[26]
	Elastic net regression feature reduction	[35]
	Khoja stemmer and Mo3jam	[37]
	Stop words removal	43]
Feature selection	Information Gain	[43]
	Correlation-based Feature Subset	[43]
Feature extraction	TF-IDF	[32], [53], [38]
	Word count/frequency	[37], [58]
	Word2Vec	[26], [58]
	Word entropy	[58]
	Words ngrams	[32]
	GloVe	[32]
	PCA and latent semantic analysis	[53]
	Bag-of-words and bag-of-concepts feature vectors	[43]
	Normalized Pointwise Mutual Information	[32]
Annotation and Sentiment Lexicons	Hedonometer happiness lexicon (LabMT)	[28], [33], [39], [40]
	ANEW lexicon	[28], [31], [33]
	PERMA lexicon	[26], [33]
	AFINN dictionary	[41]
	Arabic sentiment lexicons	[37]
	Big 5 personality	[29]
	NRC emotional lexicon	[37]
	Manual annotators	[37]
	SentiStrength V2.2	[30]
	SentiWordNet sentiment lexicon	[34]
	SentimentIntensityAnalyzer	[38]
	Pretrained models	[32]
Text analysis method	LIWC	[27], [29], [31], [33], [35]
	SCLIWC (Simplified Chinese version of LIWC)	[27]
	Valence Shift Word Graphs	[39]

TABLE 7. Employed data preparation techniques in other QoL studies.

Task	Methods	Research Works
Cleaning and Preprocessing	Min-max scaler	[52], [8]
	One hot encoding	[8]
	The inverse probability weighting	[70]
Feature selection	Information Gain	[50]
	Steepest Ascent Hill-Climbing feature selection	[56]
	XGBoost feature importance	[59]

in this table, other datatypes are presented in terms of preprocessing.

Two research works using international datasets applied preprocessing tools [8] and [52]. In [50], a feature selection technique was employed. While [56] and [59], which based their research on HRQoL, employed feature selection techniques, only the work of [70], which used a questionnaire, employed a feature selection technique. It is observed that there is a limited utilization of tools in studies pertaining to well-being.

Specifically, for these three types of data, they are already pre-processed and do not require extensive cleansing and preparation. Additionally, the practice of feature selection is infrequent, as researchers typically aim to validate hypotheses concerning the most influential indicators.

The first category of tasks, cleaning and preprocessing, includes techniques dealing with multiple aspects of data preparation, including scaling and data transformation. For international datasets, the application of techniques like min-max scaling and one-hot encoding proved to be effective in standardizing numerical features and handling categorical variables, respectively.

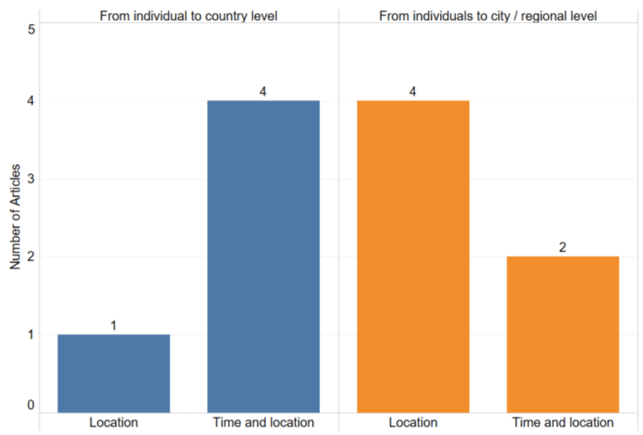

FIGURE 7. Social media works using time and location features for expanding their level of study.

These steps are crucial in ensuring that data from diverse sources can be meaningfully compared and integrated. The use of inverse probability weighting in the context of questionnaires demonstrated its potential for addressing imbalances and rare events, a key consideration in maintaining the integrity of the analysis.

The second category of tasks, feature selection, involves identifying the most relevant features or variables in the data that will be used to train the model.

This is often done through statistical analysis or by selecting a subset of the most informative and important features from machine learning algorithms. Feature selection is important because it helps to reduce the dimensionality of the data and improve the efficiency of the model.

D. MACHINE LEARNING AND STATISTICAL METHODS (RQ4)

1) MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

The prediction of wellbeing indicators holds significant importance in understanding and assessing individuals' socioeconomic status and overall standard of living. Accordingly, a lot of work has been dedicated to building predictive models for QoL using artificial intelligence. The integration of various algorithms, each selected for its unique strengths, underscores the complexity and multidimensionality of the QoL concept.

In Figure 8, we observe the most used AI algorithms for each research goal represented by the studied QoL objective. We note that machine learning algorithms were more commonly employed than deep learning algorithms. Furthermore, the most used algorithms are SVM and random forests, followed by MLP. Additionally, ANN, SVM, and random forests are the only algorithms utilized for all four dimensions.

For subjective wellbeing, researchers employed a variety of machine learning techniques, especially linear regression, NB, SVM, and RF. Multiple algorithms were utilized for studying SWB in addition to the techniques mentioned in the previous sections. The most commonly used algorithms for studying general wellbeing are SVM and SVR, followed by MLP and random forests. While SVM, LSTM, DT, and KNN, among other algorithms, were used to study HRQoL. In the evaluation of urban quality of life, knowledge representation techniques were employed instead of AI models, which is why they do not appear in this table. The specific methodologies utilized will be elaborated upon in the subsequent section.

The diversity of algorithms employed in this study reflects a thoughtful approach towards capturing the intricacies of each QoL dimension. For instance, SVM is chosen for general wellbeing and HRQoL targeted studies, potentially indicating the presence of non-linear relationships in these dimensions. This aligns with the broader understanding that these aspects of QoL can be influenced by a multitude of complex and interdependent factors.

Random forests, a versatile ensemble learning technique, find application in general wellbeing and economic wellbeing. This choice suggests an appreciation for the model's ability to capture intricate interactions within the data. Logistic regression, on the other hand, is employed for its ability to predict binary outcomes. In the context of healthcare, logistic regression can be particularly useful for predicting the likelihood of an individual experiencing a particular outcome, such as the presence or absence of an illness or disease. The inclusion of decision trees and their variants, including DT and LMT, further emphasizes the importance of classification algorithms in this context.

Time series models, such as ARIMA, FARIMA, and Holt Winters, play a crucial role in forecasting economic wellbeing indicators. This underscores the recognition of temporal patterns as pivotal factors in these particular dimensions.

Deep learning techniques, including MLP, ANNs, LSTM, RNN, and CNN, are strategically employed for their capability to capture complex relationships. Their application in assessing general wellbeing attests to the acknowledgment of the complex interplay of factors in these dimensions.

The consideration of principal component regression, partial least squares regression, MARS, Gaussian process, elastic net, and decision table, though less prevalent, implies their specific utility in addressing unique characteristics or patterns within the data.

Unsupervised Learning approaches, although less prevalent, offer valuable insights into latent patterns within QoL without the aid of labeled outcomes or predefined target variables. Among such methods, LDA stands out in topic modeling, a technique pivotal in identifying and categorizing abstract topics within large text corpora.

While the emphasis in this part is on predictive models, it is imperative to acknowledge the application of unsupervised methods in pre-processing phases, as delineated in (RQ3). This encompasses the utilization of sentiment lexicons, text analysis methodologies, and a variety of feature extraction techniques.

Type of algorithm	Type of approach	Used Algorithms	Subjective wellbeing	General wellbeing	HRQoL	Economic wellbeing	Total
Machine learning	Supervised	SVM/SVR	2	5	5	1	13
		Random Forests	2	3	1	4	10
		Logistic regression	1	1	2		4
		DT		2	2		4
		Naive Bayes	2	1			3
		Linear regression	3				3
		Lasso Regression	1	1		1	3
		Gradient boosting		1		2	3
		XGBoost		1	1		2
		Time series models				2	2
		KNN			2		2
		Ridge regression				1	1
		Principal Component Regression				1	1
		Partial least squares regression				1	1
		MARS				1	1
		Logistic model tree		1			1
		Gaussian Process				1	1
		FARIMA				1	1
		Elastic Net				1	1
		Decision Table				1	1
		ARIMA				1	1
		AdaBoost Regressor		1			1
	Unsupervised	LDA classifier	3				3
Deep learning	Supervised	MLP	1	4		1	6
		ANN	1	1	1	1	4
		LSTM			2		2
		RNN	1				1
		CNN			1		1
Grand Total			17	22	17	21	77

FIGURE 8. **The algorithms used for predicting QoL by type of algorithm, approach, and QoL dimension.**

Together, supervised and unsupervised learning methodologies provide a holistic approach to understanding and enhancing QoL across various domains.

2) PERFORMANCE EVALUATION

In order to assess the performance of the AI algorithms they proposed, researchers employed a variety of metrics. Figure 9 shows the metrics that were most commonly used in the studied research.

Among classification models, accuracy was the most frequently used metric for predicting quality of life, followed by F-score and precision.

For evaluating regression models, RMSE was used in nine papers, R-squared and MAE were used in five and four papers, respectively, and relative absolute error (RAE) was used in three papers.

These metrics were used to evaluate the performance of the proposed machine learning models. Table 8 shows the performances of the proposed models for each QoL objective. Each row in the table represents a different research work, and the columns provide information about the data used, the features selected, the target variable, the best machine learning model, the evaluation metric used, and the best value achieved by that model.

For health-related QoL, the KNN model proposed in the work [59] achieved the best R-squared score of 0,940 for predicting HRQoL scores in a rare disease. This work used XGBoost feature importance as feature selection. This is the highest performance using this metric in the examined literature, additionally, this model achieved an RAE of 0.25.

Further, machine learning algorithms achieved the highest accuracy and precision of 95.4% and 95.6%, respectively. These results were obtained in the work of [49], which focused on predicting depression using machine learning algorithms. This work used an ensemble classifier composed of ANN, SVM, KNN, and DT. While the work of [60] achieved the minimum RMSE of 0.2580 using a fuzzy neural network using HRQoL survey data.

While for predicting general wellbeing, the best accuracy is 96.42%, which is attained by [52] the ensemble model of DT, RF, and SVR based on data from the BLI, this work used the Min Max scaler during the data preparation phase. The second-highest r-squared in the literature and the best for this datatype is 0.907, reached by [72] for studying the BLI using

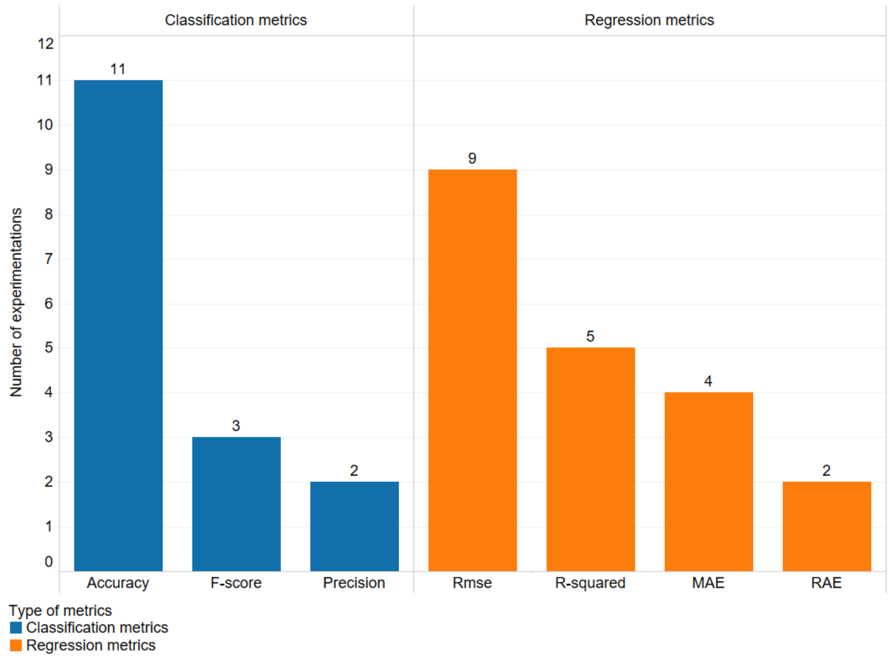

FIGURE 9. Most used performance evaluation metrics to evaluate the proposed models.

regression analysis, and the minimum RMSE of 0.0656 is achieved by [8] using LASSO regression for predicting the World Happiness Index, noting that this work used two data preparation techniques, namely, one hot encoding and a min max scaler.

In addition, for economic wellbeing, the minimal MAE found in the literature is achieved in the work [64], in which a model to forecast and study determinants of GDP is proposed.

The subjective well-being study is highly dependent on social media data, as we have presented in the previous sections. ML algorithms achieved an accuracy of 91.3% in the work of [42] using LDA for studying Twitter data and Bayesian net during the data preparation phase; this work also achieved the best f-score in the literature of 0.9129. Meanwhile, [28] achieved the best RMSE for applying linear regression to predict well-being based on the ANEW lexicon for annotation. And [29] achieved the best MAE score = 0.67 for SWB based on the Big 5 personality features and RF. Additionally, for studying student SWB, the work of [47] achieved the highest R-squared of 0.701 by using ANN on survey data.

In the study of diverse QoL research works, used tools and algorithms, datasets, and their associated features, it becomes evident that the choice of features and modeling techniques is tailored to the specific characteristics of the data and research objectives.

The selection of models, ranging from traditional linear regression to sophisticated ensemble methods like random forests and gradient boosting, highlights the need to capture and understand complex relationships within the data in order to enhance accuracy.

In contrast, simpler models such as logistic regression are preferred when exploring straightforward relationships, as seen in physical and mental health analyses. The employment of time series models like FARIMA for economic and unemployment data showcases the importance of temporal trends in these domains.

Additionally, the incorporation of text analysis techniques, including topic modeling using LDA and sentiment analysis using models like LASSO, underscores the growing importance of extracting valuable insights from unstructured text data, particularly from social media platforms like Twitter and Facebook.

These outcomes collectively demonstrate the adaptability of data analysis techniques to diverse datasets, leading to a deeper understanding of quality of life, health, and socio-economic indicators in various contexts, ultimately informing evidence-based decisions and policies.

Overall, the table above provides a summary of different works and the machine learning models that performed best in each specific task related to investigating wellbeing dimensions, along with the evaluation metrics used and the values achieved. It can be useful for comparing different models and approaches across different datasets and target variables.

3) STATISTICAL ANALYSIS TOOLS

Examining the results obtained from the measurement of well-being indicators is a valuable approach, particularly for policymakers, as the use of analytical tools can provide significant insights. These tools can help identify the key factors influencing the standard of living of the studied samples.

TABLE 8. Performance evaluation of proposed models for each QoL research goal.

QoL dimension	Ref	Dataset	Features	Best model	Metric	Best value
HRQoL	[60]	WHOQoL-100 survey	Demographic, Economic, Health, Social, and Community	Fuzzy Neural Network	**RMSE**	**0.258**
	[49]	NHANES	Mental health	Ensemble binary classifier combining ANN, SVM, KNN, and DT	**F- score** **Accuracy** **Precision**	**0.976** **95.4%** **95.6%**
	[44]	NHANES	Physical and mental health	Stepwise logistic regression	Accuracy F-score	93% 0.49
	[59]	The ApreciseKUre database	Physical and mental health	KNN	**R squared** **RAE**	**0.94** **0.25**
	[56]	Smartphone accelerometer and gyroscope sensors data	Physical and mental health, and meta information	SVM regression	MAE RAE	0.117 0.519
Economic wellbeing	[64]	World bank data	World Development Indicators, including GDP per capita	Linear regression	**MAE** **RAE**	**0.01801** **0.00982**
	[75]	GDP per capita	Multiple economic indicators related to GDP	Gradient boosting regressor	RMSE MAPE	0.35 19.86
	[54]	Eurostat database	Unemployment and Jobs	FARIMA	Average RMSE Average MAE	0.163 0.130
General wellbeing	[45]	Survey data on multiple aspects	Demographic, educational, health-related, and social indicators	Stacking-based model combining LMT and sequential minimal optimization (SVM)	Accuracy	84.20
	[52]	OECD's Better Life Index data	BLI indicators	Ensemble model of DT, RF, and SVR	**Accuracy** RMSE	**96.42%** 0.29
	[8]	World Happiness Index data	WHI indicators	LASSO	**RMSE** R squared	**0.0656** 0.8954
	[72]	OECD's Better life index and Eurostat datasets	BLI indicators, GDP per capita, Gini coefficients, Gender Gap Index	Regression analysis	**R squared** Adjusted R2	**0.907694** 0.905456
Subjective Wellbeing	[42]	Twitter	Tweets, topics, events and meta information including time, location, and author information	LDA classifier	**Accuracy** **F-score**	**91.30%** **0.9129**
	[28]	Twitter	Tweets and meta information including locations	Linear regression	MAE **RMSE**	0.92 **1.22**
	[36]	Twitter	Tweets and meta information including locations	RNN	Accuracy	85.4%
	[32]	Twitter	Tweets, meta information, and academic discourse	Logistic regression	**Precision** Recall F-score	**0.71** 0.69 0.64
	[29]	Facebook and myPersonality project dataset	Big 5 personality features, and user information	RF	MAE	**0.67**
	[47]	Survey on College students	Demographic, economic, education, health, services, social and community survey	ANN	**R squared**	**0.701**

In research, various techniques were employed to analyze quality of life indicators.

Figure 10 categorizes the used statistical analysis tools into different types based on their purpose, including correlation and descriptive statistics, statistical tests, regression analysis, structural models, inferential statistics, analysis of variance, factor analysis, and other tools.

Each tool is useful in different scenarios and research questions, and selecting the appropriate tool depends on the nature of the data and the research question. It is important to have a good understanding of these tools to conduct effective statistical analysis and draw valid conclusions from the data.

The most frequently used technique is correlation measurement tools in 23 research works, especially Pearson's correlation, which was used in 15 works and applied to QoL dimensions. Correlation analysis is a statistical technique that measures the strength and direction of the relationship between two or more variables. This approach is commonly used in research to explore the extent to which two variables are related. The popularity of correlation analysis in research reflects its utility in identifying patterns and relationships in data, which can aid in drawing conclusions and informing decision-making processes.

Additionally, analysis of variance was employed in seven research works. By analyzing the variance in the data across the different groups, ANOVA can determine whether there is a statistically significant difference between the means of the groups. Further, regression analysis was also one of the most used tools, particularly for questionnaires and international datasets. It is mainly used to model the relationship between two or more variables.

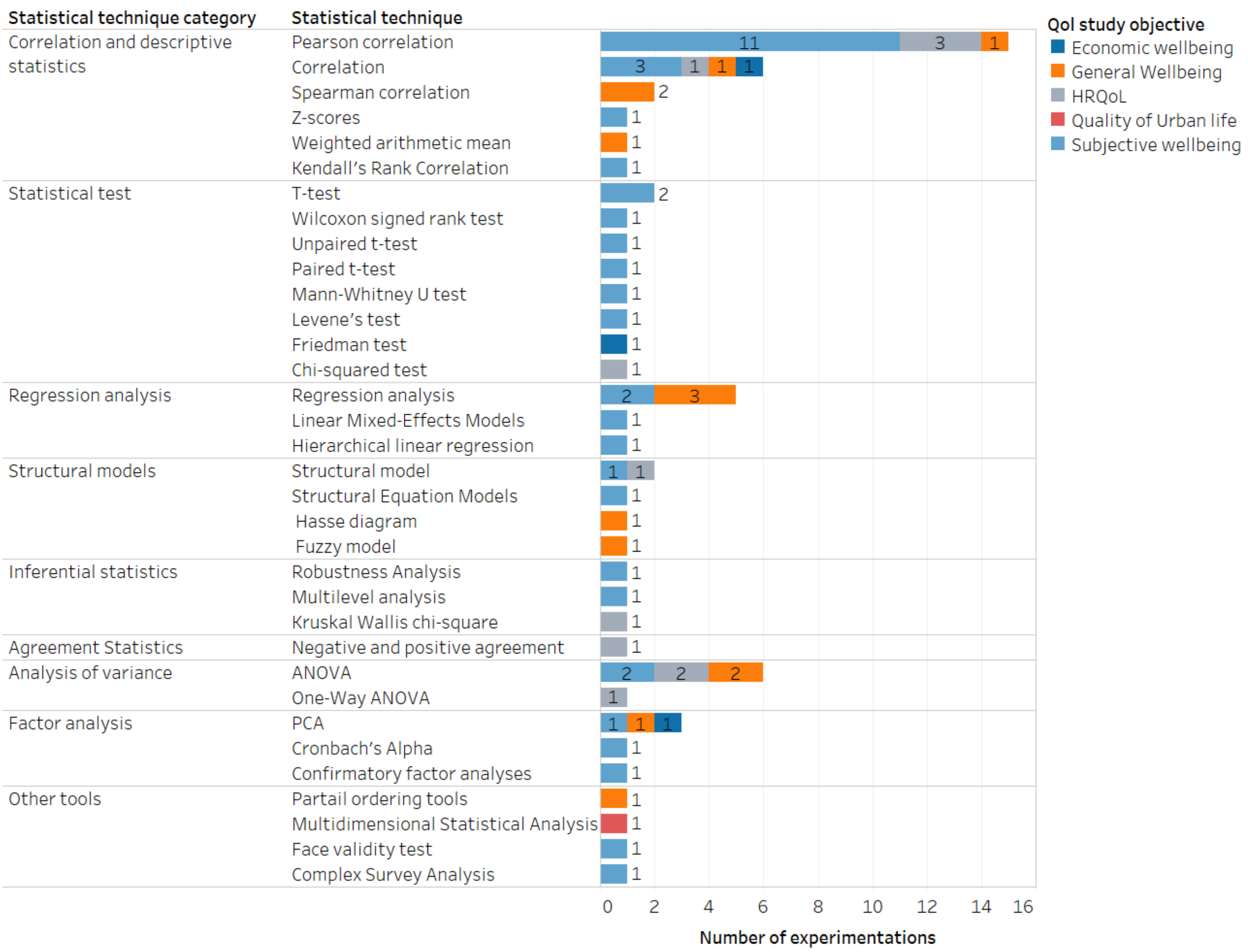

FIGURE 10. The statistical techniques used for analyzing QoL data by type and datatype.

In studies focused on subjective well-being, several key statistical techniques have yielded significant insights. We note that statistical techniques are mostly applied to this datatype, which makes it the type with the most preprocessing techniques and the most statistical techniques. Correlation measures, especially Pearson Correlation played pivotal roles in identifying strong associations between various factors and individuals' self-reported wellbeing. It is worth emphasizing that correlation techniques were extensively utilized in studies pertaining to subjective well-being. This is primarily due to researchers often augmenting social media data with additional datasets to explore correlations with various events and phenomena or other metrics. For instance, the analysis of the relationship between happiness and gold prices, as previously discussed in [40].

For general well-being, correlation and descriptive statistics played a significant role in identifying key contributing factors. Regression analysis proved pivotal in understanding the collective influence of economic wellbeing, the urban environment, and societal pressures. These approaches collectively enriched the comprehension of the diverse factors influencing general wellbeing, which is the second studied QoL dimension after SWB from the perspective of statistical tools.

Finally, the judicious selection of statistical techniques tailored to the unique characteristics of each dimension of quality of life has yielded substantial insights. These techniques have empowered researchers to quantify and understand the relationships between various factors and the overall wellbeing within each respective domain. From uncovering subjective perceptions to analyzing objective indicators, these methods have been instrumental in gaining a comprehensive understanding of quality of life from multiple perspectives.

E. QOL MODELING (RQ5)

Modeling and building ontology-driven models of QoL indicators is a commonly explored challenge in the literature. Researchers have sought to model or propose indexes using various methodologies, such as assembling multiple concepts to form a new ontology-based index of well-being. Ontology-driven models offer a systematic way to categorize and integrate various dimensions of well-being, allowing for a more nuanced analysis and comparison across different

TABLE 9. Research works on QoL modeling by QoL objective, indicators and level of study.

QoL study objective	Level of study	Qol indicators	Ref
General Wellbeing	Country	Better Life Index	[91]
		Country level QoL	[80]
		The Global Reporting Initiative sustainability indicator (GRI), BLI, and UN Social Indicators	[81]
	General	QoL indicators as a concept	[85], [88]
HRQoL	General	HRQOL indicators	[94]
QoUL	City	Economic, societal, and environmental development indicators	[92]
		ISO 37120	[84], [89]
		ISO 37121 Sustainable development	[83]
		ISO 37120 educational indicators	[86], [93]
		311 city services	[82]
		Transportation indicators	[87], [90]

TABLE 10. An overview of the used tools to model QoL indicators.

Type of technique	Used techniques	Research works
Modeling approach	Ontology	[81], [82], [83], [84], [86], [85], [87], [88], [89], [90], [92], [93]
	Fuzzy model	[91]
	METHONTOLOGY	[81]
	MIMIC approach	[80]
	Data management system	[94]
Ontology building tool	Prolog	[87]
	Protégé	[81], [82], [87], [92], [93]
	WebProtégé	[83], [86]
Semantic web language	OWL/OWL2	[82], [86], [84], [87], [92]
	RDF	[92], [93]
	SPARQL	[82], [87], [92], [93]

contexts. These models are particularly valuable in their ability to assemble multiple concepts into coherent frameworks that reflect the complexity of well-being.

The research works, which are illustrated in Table 9, propose semantic models of quality of life at various levels, especially city and country levels. The majority of research on ontologies in the studied literature focused on the city level, and the majority of studies used the ISO 37120 index as a QoUL indicator.

Table 10 shows the works that proposed models of QoL using a set of tools rather than just discussing the targeted concept and its different aspects. For this aim, researchers mainly used ontology. It was used to create a framework for storing and analyzing data, as well as to explore the underlying assumptions and relationships between concepts. Ontology is crucial for modeling quality of life indicators due to its multidimensional nature.

Most of the works used ontologies; while other works used other approaches, like in the work [91] which used fuzzy models and decision-making techniques, [80] used the MIMIC approach, which is a statistical modeling technique that is used to explore the relationships between multiple causes and multiple indicators. Further, [94] proposed a model for storing data based on cutting-edge technologies such as IoT for collecting HRQoL data, cloud storage, and blockchain for storing data in a secure way. Additionally, OWL is the most widely used semantic web language, in addition to SPARQL. The most commonly used tool for building ontologies is Protégé. Additionally, the implementation of the ISO 37120 standard allowed researchers to study and analyze data from the cities that implemented it [77].

V. DISCUSSION

The review of prior literature highlights important contributions by researchers to measuring and predicting QoL. In this context, most of the studies proposed artificial intelligence models to achieve these goals. However, there is still work to be accomplished to fill gaps and overcome limitations in the literature. In this section, we will summarize our findings, discuss the limitations of the studied works, and make suggestions for future research.

Through this SLR, we have tried to outline the different proposed approaches to quantifying and analyzing well-being. The QoL concept can be related to multiple fields and levels. To reach their objectives, researchers used a variety of QoL indicators that can be categorized according to the studied aspect of well-being and the level of the study.

Three levels were identified: individual, city, and country. The first level is related to the study of happiness for a sample of individuals. This level is largely related to HRQoL and SWB. City level is strongly related to the quality of urban life and, in some cases, SWB. This latter is measured at this level using large samples of individuals with geotagged data extracted from social networks. At last, country level is related to assessing QoL at the national and international scales.

Researchers addressed quality of life issues by proposing several approaches. First, they were able to quantify and predict the quality of life through measures that made the studied samples comparable, depending on the level of study in question, by proposing predictive models. Secondly, they analyzed QoL indicator measures to solve problems, test hypotheses, and answer research questions related to QoL domains. This enabled accurate assessment of well-being and relevant analysis based on cutting-edge algorithms and statistical methods, which can provide decision-makers with the right vision to make the necessary decisions and propose countermeasure solutions. Thirdly, the use of semantic web tools enabled the development of universal ontological models for indicators, which help in storing and sharing data between different entities in a reliable and consistent way. Figure 11 shows how these approaches are related to multiple AI concepts.

For instance, machine learning and deep learning algorithms were used by researchers for both predictive and analytic purposes to build models able to measure, predict, and analyze QoL indicators. Furthermore, several studies used NLP techniques and AI algorithms to propose

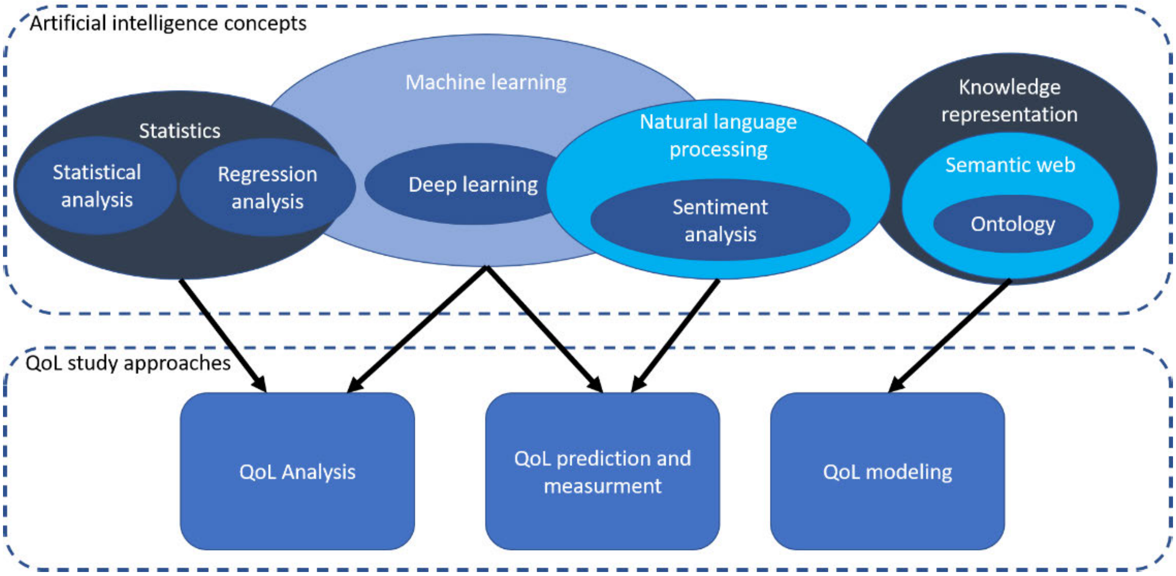

FIGURE 11. Artificial intelligence concepts and their use for quality-of-life studies.

sentiment analysis methods capable of interpreting text and exploring feelings and opinion polarity, particularly for investigating social media data to assess subjective well-being. Moreover, regression analysis and statistical tools were used to build multipurpose analysis models to investigate the ups and downs, most influential factors, and determinants of well-being indicators. Additionally, knowledge representation concepts were used in this context to model QoL indicators, especially semantic approaches.

To summarize, researchers used multiple approaches, which we categorize in this figure as: 1) sentimental analysis approaches; 2) machine learning models; 3) statistical approaches; and 4) semantic approaches, based on the AI concepts.

Additionally, multiple types of data were used to achieve research objectives. These datatypes can be categorized into four groups, which are: social media, questionnaires, international datasets, and HRQoL data, which includes clinical, EMR, and sensor data. Social media data and questionnaires are the most used in data-driven research works compared to international and HRQoL. We emphasize the diversity of the employed data sources, to which the researchers contributed. In many research works, authors had to build questionnaires with specific questions to assess the targeted QoL indicators in their research, notably when measuring SWB and HRQoL.

While in some works, multiple types of data were used when primary data missed important aspects to achieve research objectives, researchers were forced to use additional datasets and investigate correlations to overcome the data's limitations and test hypotheses in order to achieve research goals. In summary, studying quality of life using various data types is a multidimensional endeavor that requires creativity, adaptability, and a diverse set of data collection approaches. Researchers in this field must navigate the absence of standard datasets by employing methods such as web scraping, extracting data from reports and available APIs, and considering qualitative data to gain a holistic understanding of well-being. Each data type has its advantages and limitations, and the choice of method depends on the research questions and objectives at hand. However, it is essential to consider the potential limitations and biases inherent in each use case when drawing conclusions or making policy recommendations.

In this review, we concluded that QoL indicators can be categorized into: 1) economic well-being, which is directly related to economic measures such as GDP and can be based on data from multiple international databases such as the World Bank, OECD websites, and the WHI report; 2) health-related quality of life, which is defined as a couple of well-being measures related to a specific disease or health condition and is mainly related to assessing the health status of individuals using HRQoL data; 3) subjective well-being, which is a measure of happiness, satisfaction with life, and subjective components of well-being, is mainly measured based on social media data by applying sentiment analysis and text mining approaches; 4) Quality of urban life is an assessment of living standards in a given city based on well-defined indicators of QoL and city services; and 5) general well-being, which can be related to multiple levels, is a combination of multiple aspects of quality of life used to form an index.

The primary contribution of researchers in the QoL study is building sentimental analysis models using cutting-edge NLP methods. These models allow the measurement of the subjective well-being of groups of people based on thousands to millions of social media posts from Twitter, Facebook, and Sina Weibo. In this context, researchers faced many challenges related to the non-standardization of languages on social platforms. Several NLP methods, including LIWC for text analysis and LDA for topic modeling, were used to address these issues.

Researchers conducted several experiments aiming to assess standards of living using a wide variety of ML and DL algorithms such as RF, SVM, and ANN, which are the most used in the literature and the best performing ones, including KNN, Lasso regression, linear regression, and logistic regression. We noted that deep learning algorithms were not widely used due to the small size of QoL data in some contexts. In this case, data augmentation can be applied and interpreted before applying DL algorithms, which can also increase the performance of the predictive models.

The current landscape of machine learning within the realm of QoL assessment suggests a significant reliance on supervised learning for its predictive strength. While this has yielded many insights, there remains a considerable gap in utilizing unsupervised learning to its full potential. Unsupervised learning's power to discover hidden patterns without pre-labeled outcomes could unlock new dimensions of QoL that are not apparent through supervised methods alone. Methods such as clustering, association rule mining, and anomaly detection could identify unexpected correlations and trends that might escape traditional supervised models.

Additionally, various statistical analysis techniques were used to study and analyze well-being indicators, which yielded rich insights, most notably regression analysis and Pearson correlation. Another contribution is modeling quality of life indicators while taking advantage of semantic web technologies, especially OWL, SPARQL, and Protégé. Specifically at the city level, these languages and tools were used to propose semantic models based on QuoL, primarily on ISO 37120.

However, these models and techniques were applied to a limited number of quality-of-life indicators, especially at the country level, which are: the world happiness index, the better life index, the human development index, gross national happiness, and GDP. While many other indicators were not investigated, such as the Happy Planet Index, Genuine Progress Indicator, Where-to-Be-Born Index, and Legatum Prosperity Index, a deep study must be conducted to reveal the reason behind the lack of research on these indicators.

Furthermore, after examining quality of life indicators projected on those levels, we discovered that studying QoL from these perspectives can be sufficient for meeting various development and sustainability goals on all three levels. Nevertheless, the research on quality of life did not address the level of households, which is a very crucial level. More research on household QoL can provide important insights, considering households as important components of every society. Hence, the Demographic and Health Surveys (DHS) proposed an indicator named the DHS Wealth Index [99]. Using this index, artificial intelligence techniques can be used to reveal multiple aspects of well-being at this level of study.

Despite the fact that several sentiment analysis models have been proposed to measure subjective well-being indicators, no global quality of life index uses this technology as an official measurement tool for SWB. These proposed models have to be considered by policymakers and international organizations that publish QoL indicators, especially those that rely on survey data, since social media can offer more representative information on millions of users than surveys based on a limited number of respondents. Notable advantages of NLP models over questionnaires include the ease of their implementation and the ability to reach a large sample of participants. Traditional surveys and questionnaires often require significant time and resources to design, distribute, and collect responses. Conversely, NLP models can autonomously process big data from various sources, including social media, APIs, forums, and online reviews, offering a faster and more scalable approach to gathering insights about individuals' experiences and perceptions at larger scales. Moreover, one of the most interesting approaches found in the literature is expanding the level of study from the publications of individuals to study happiness at the city and country levels. Further, the advanced NLP capabilities of Large Language Models (LLMs) can enable the nuanced analysis of vast amounts of textual data, from social media sentiment to qualitative interviews, offering new insights into human well-being. This potential makes LLMs a valuable tool for exploring the complex dimensions of QoL, identifying patterns and trends in public perceptions, attitudes, and experiences. This could significantly enhance our understanding of well-being, informing more effective policies and interventions. This approach aligns with the evolving needs of interdisciplinary research in capturing the multifaceted nature of human life quality.

Lastly, a semantic model of quality of life, especially at the country level, should be developed similarly to the city-level model. Therefore, the efforts of international organizations should be unified to build a universal model that describes quality of life components in the best possible way. This will make the analysis and prediction of well-being easier and can help policymakers and organizations interested in enhancing the welfare and quality of life of populations.

In QoL research, the impact of practical implications often surpasses that of conceptual implications. While conceptual implications are crucial for advancing theoretical understanding and shaping the academic discourse surrounding well-being and life satisfaction, practical implications directly influence the lives of individuals and communities. The practical insights gleaned from such research have the power to inform policies, interventions, and programs aimed at enhancing the overall well-being of populations. By offering data-based recommendations and actionable strategies, quality of life research with applied AI models can lead to immediate improvements in areas such as healthcare, education, urban planning, and social services. Therefore, in this context, the real-world impact of practical implications stands as a cornerstone in the pursuit of enhancing the well-being of society.

VI. CONCLUSION

In order to raise the standards of living for people worldwide, a fair assessment of their mental and material well-being must be realized. Data mining tools and the vision of data scientists are required for the attainment of this goal. In this paper, we examined the research works to identify the various AI approaches used in the literature to investigate the concept of QoL.

This SLR is an exhaustive work that can help researchers in the QoL domain. A deep analysis of objectives, approaches, datasets, and indicators is provided. Therefore, we concluded that data scientists may provide insights about the most important indicators through sentimental analysis, machine learning algorithms, and semantic tools. Proposed models can investigate affecting factors, explain ups and downs, and estimate future outcomes of QoL indicators at various levels. Predictive ML models can assist policymakers and international organizations in making the best decisions to ensure that people have a better standard of living. More importantly, the use of sentiment analysis based on NLP and ML has enabled the assessment of subjective well-being directly from the large amount of social media data in the connected world. We have systematically examined effective methodologies and significant scholarly contributions. Additionally, we have outlined the existing research gaps pertaining to a spectrum of factors, including used datasets, algorithms and methods, and QoL indicators and levels. Subsequently, we presented deficiencies in research and proposed possible solutions.

For future research, we aim to examine how the indicators and their sub-indicators relate to each other in order to create a comprehensive model of general well-being. Additionally, we plan to use more social media information to gather and structure details for creating models basing on cutting-edge methodologies such as LLMs for studying happiness using NLP and ontologies.

Moreover, universal models can help standardize the assessment of SWB, thus making it comparable. Following that, we aim to suggest a detailed model of well-being that covers individuals, cities, and countries levels from multiple perspectives. Additionally, exploring how text mining and computer vision tools can help us investigate overlooked aspects of quality of life and create new sources of data.

VII. LIMITATIONS

In order to comprehend the scope and general applicability of the results, it is crucial to recognize the limitations of the current study. The study did not include projections on the problematic aspects of the Internet of Things. Other researchers may offer alternative viewpoints regarding the challenges and opportunities associated with IoT beyond those considered in this study. Moreover, since concluding data collection, advancements may have occurred in pertinent fields, notably in areas such as LLMs. Recent research or methodologies introduced post-2022 might offer supplementary insights or perspectives not addressed in this study. Furthermore, language barriers may have resulted in the exclusion of non-English publications, which might introduce bias into the analysis in some contexts. Although these limitations do not invalidate the results of the study, they must be acknowledged while interpreting the findings and evaluating their relevance to other populations and contexts.

REFERENCES

[1] R. Gillingham and W. S. Reece, "Analytical problems in the measurement of the quality of life," *Social Indicators Res.*, vol. 7, nos. 1–4, pp. 91–101, Jan. 1980, doi: 10.1007/bf00305594.

[2] T. J. VanderWeele, C. Trudel-Fitzgerald, P. Allin, C. Farrelly, G. Fletcher, D. E. Frederick, J. Hall, J. F. Helliwell, E. S. Kim, W. A. Lauinger, M. T. Lee, S. Lyubomirsky, S. Margolis, E. McNeely, N. Messer, L. Tay, V. Viswanath, D. Węziak-Białowolska, and L. D. Kubzansky, "Current recommendations on the selection of measures for well-being," *Preventive Med.*, vol. 133, Apr. 2020, Art. no. 106004, doi: 10.1016/j.ypmed.2020.106004.

[3] I. Bache, "Wellbeing," in *Evidence, Policy and Wellbeing* (Wellbeing in Politics and Policy), I. Bache, Ed. Cham, Switzerland: Springer, 2020, pp. 29–51, doi: 10.1007/978-3-030-21376-3_3.

[4] E. Diener, S. Oishi, and L. Tay, "Advances in subjective well-being research," *Nature Hum. Behav.*, vol. 2, no. 4, pp. 253–260, Feb. 2018, doi: 10.1038/s41562-018-0307-6.

[5] P. Dolan and R. Metcalfe, "Measuring subjective wellbeing: Recommendations on measures for use by national governments," *J. Social Policy*, vol. 41, no. 2, pp. 409–427, Apr. 2012, doi: 10.1017/s0047279411000833.

[6] A. T. Jebb, L. Tay, E. Diener, and S. Oishi, "Happiness, income satiation and turning points around the world," *Nature Hum. Behav.*, vol. 2, no. 1, pp. 33–38, Jan. 2018, doi: 10.1038/s41562-017-0277-0.

[7] A. Jannani, N. Sael, and F. Benabbou, "Machine learning for the analysis of quality of life using the world happiness index and human development indicators," *Math. Model. Comput.*, vol. 10, no. 2, pp. 534–546, 2023, doi: 10.23939/mmc2023.02.534.

[8] A. Jannani, N. Sael, and F. Benabbou, "Predicting quality of life using machine learning: Case of world happiness index," in *Proc. 4th Int. Symp. Adv. Electr. Commun. Technol. (ISAECT)*, Dec. 2021, pp. 1–6, doi: 10.1109/ISAECT53699.2021.9668429.

[9] M. Romero, D. Vivas-Consuelo, and N. Alvis-Guzman, "Is health related quality of life (HRQoL) a valid indicator for health systems evaluation?" *SpringerPlus*, vol. 2, no. 1, p. 664, Dec. 2013, doi: 10.1186/2193-1801-2-664.

[10] S. Atkinson, A.-M. Bagnall, R. Corcoran, J. South, and S. Curtis, "Being well together: Individual subjective and community wellbeing," *J. Happiness Stud.*, vol. 21, no. 5, pp. 1903–1921, Jun. 2020, doi: 10.1007/s10902-019-00146-2.

[11] C. C. Walther, "From individual wellbeing to collective welfare," *Humanitarian Leader*, Feb. 2020, doi: 10.21153/thl2020volno0art1014.

[12] *Quality of Urban Life IG*. Accessed: Feb. 20, 2022. [Online]. Available: https://www.rd-alliance.org/groups/urban-quality-life-indicators.html

[13] A. A. Arman, A. E. Abbas, and R. Hurriyati, "Analysis of smart city technology initiatives for city manager to improve city services and quality of life based on ISO 37120," in *Proc. 2nd Int. Conf. Electron. Governance Open Soc., Challenges Eurasia*. New York, NY, USA: Association for Computing Machinery, Nov. 2015, pp. 193–198, doi: 10.1145/2846012.2846025.

[14] J. Hendler and F. van Harmelen, "The semantic web: Webizing knowledge representation," in *Handbook of Knowledge Representation*, vol. 3, F. van Harmelen, V. Lifschitz, and B. Porter, Eds. Amsterdam, The Netherlands: Elsevier, 2008, ch. 2, pp. 821–839.

[15] S. Singh and P. Kaur, "Social media trends and prediction of subjective well-being: A literature review," in *Proc. ICCCCS*, 2014, pp. 741–751, doi: 10.1007/978-981-10-3770-2_70.

[16] M. Luhmann, "Using big data to study subjective well-being," *Current Opinion Behav. Sci.*, vol. 18, pp. 28–33, Dec. 2017, doi: 10.1016/j.cobeha.2017.07.006.

[17] E. Pleeging, M. Burger, and J. van Exel, "The relations between hope and subjective well-being: A literature overview and empirical analysis," *Appl. Res. Quality Life*, vol. 16, no. 3, pp. 1019–1041, Jun. 2021, doi: 10.1007/s11482-019-09802-4.

[18] S.-H. Liao, P.-H. Chu, and P.-Y. Hsiao, "Data mining techniques and applications—A decade review from 2000 to 2011," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11303–11311, Sep. 2012, doi: 10.1016/j.eswa.2012.02.063.

[19] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2016, pp. 1310–1315.

[20] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–36, Sep. 2018, doi: 10.1145/3234150.

[21] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019, doi: 10.1109/ACCESS.2019.2912200.

[22] K. R. Chowdhary, "Natural language processing," in *Fundamentals of Artificial Intelligence*, K. R. Chowdhary, Ed. New Delhi, India: Springer, 2020, pp. 603–649, doi: 10.1007/978-81-322-3972-7_19.

[23] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of Twitter data," in *Proc. Workshop Lang. Social Media*, Portland, OR, USA, Jun. 2011, pp. 30–38. Accessed: Oct. 7, 2023. [Online]. Available: https://aclanthology.org/W11-0705

[24] D. M. E.-D.-M. Hussein, "A survey on sentiment analysis challenges," *J. King Saud Univ.-Eng. Sci.*, vol. 30, no. 4, pp. 330–338, Oct. 2018, doi: 10.1016/j.jksues.2016.04.002.

[25] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.

[26] C. Kuang, Z. Liu, M. Sun, F. Yu, and P. Ma, "Quantifying Chinese happiness via large-scale microblogging data," in *Proc. 11th Web Inf. Syst. Appl. Conf.* Tianjin, China: IEEE, Sep. 2014, pp. 227–230, doi: 10.1109/WISA.2014.48.

[27] B. Hao, L. Li, R. Gao, A. Li, and T. Zhu, "Sensing subjective well-being from social media," in *Active Media Technology* (Lecture Notes in Computer Science), vol. 8610, G. Schaefer, S. T. Vuong, and Y.-S. Kim, Eds. Cham, Switzerland: Springer, 2014, pp. 324–335, doi: 10.1007/978-3-319-09912-5_27.

[28] J. Loff, M. Reis, and B. Martins, "Predicting well-being with geo-referenced data collected from social media platforms," in *Proc. 30th Annu. ACM Symp. Appl. Comput.* Salamanca, Spain: ACM, Apr. 2015, pp. 1167–1173, doi: 10.1145/2695664.2695939.

[29] S. Collins, Y. Sun, M. Kosinski, D. Stillwell, and N. Markuzon, "Are you satisfied with life? Predicting satisfaction with life from Facebook," in *Social Computing, Behavioral-Cultural Modeling, and Prediction* (Lecture Notes in Computer Science), vol. 9021, N. Agarwal, K. Xu, and N. Osgood, Eds. Cham, Switzerland: Springer, 2015, pp. 24–33, doi: 10.1007/978-3-319-16268-3_3.

[30] A. O. Durahim and M. Coşkun, "#iamhappybecause: Gross national happiness through Twitter analysis and big data," *Technol. Forecasting Social Change*, vol. 99, pp. 92–105, Oct. 2015, doi: 10.1016/j.techfore.2015.06.035.

[31] M. E. Larsen, T. W. Boonstra, P. J. Batterham, B. O'Dea, C. Paris, and H. Christensen, "We feel: Mapping emotion on Twitter," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 4, pp. 1246–1252, Jul. 2015, doi: 10.1109/JBHI.2015.2403839.

[32] S. Volkova, K. Han, and C. Corley, "Using social media to measure student wellbeing: A large-scale study of emotional response in academic discourse," in *Social Informatics* (Lecture Notes in Computer Science), vol. 10046, E. Spiro and Y.-Y. Ahn, Eds. Cham, Switzerland: Springer, 2016, pp. 510–526, doi: 10.1007/978-3-319-47880-7_32.

[33] C. Yang and P. Srinivasan, "Life satisfaction and the pursuit of happiness on Twitter," *PLoS ONE*, vol. 11, no. 3, Mar. 2016, Art. no. e0150881, doi: 10.1371/journal.pone.0150881.

[34] K. Singh, H. K. Shakya, and B. Biswas, "Happiness index in social network," in *Advanced Informatics for Computing Research* (Communications in Computer and Information Science), vol. 712, D. Singh, B. Raman, A. K. Luhach, and P. Lingras, Eds. Singapore: Springer, 2017, pp. 261–270, doi: 10.1007/978-981-10-5780-9_24.

[35] L. Chen, T. Gong, M. Kosinski, D. Stillwell, and R. L. Davidson, "Building a profile of subjective well-being for social media users," *PLoS ONE*, vol. 12, no. 11, Nov. 2017, Art. no. e0187278, doi: 10.1371/journal.pone.0187278.

[36] B. Pauken, M. Pradyumn, and N. Tabrizi, "Tracking happiness of different U.S. cities from tweets," in *Big Data—BigData 2018* (Lecture Notes in Computer Science), vol. 2018, F. Y. L. Chin, C. L. P. Chen, L. Khan, K. Lee, and L.-J. Zhang, Eds. Cham, Switzerland: Springer, 2018, pp. 140–148, doi: 10.1007/978-3-319-94301-5_11.

[37] A. A. Alharbi, H. A. Alotebii, and A. A. AlMansour, "Towards measuring happiness in Saudi Arabia based on tweets: A research proposal," in *Proc. 1st Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS)*, Apr. 2018, pp. 1–4, doi: 10.1109/CAIS.2018.8442024.

[38] A.-L. Alten, G. Gadre, S. Kulkarni, and C.-S. Wu, "Analyzing happiness index based on geographical locations," in *Proc. 2nd Int. Conf. Artif. Intell. Big Data (ICAIBD)*. Chengdu, China: IEEE, May 2019, pp. 45–51, doi: 10.1109/ICAIBD.2019.8837010.

[39] A. Al Shehhi, J. Thomas, R. Welsch, I. Grey, and Z. Aung, "Arabia Felix 2.0: A cross-linguistic Twitter analysis of happiness patterns in the united Arab Emirates," *J. Big Data*, vol. 6, no. 1, p. 33, Apr. 2019, doi: 10.1186/s40537-019-0195-2.

[40] H. Byström, "Happiness and gold prices," *Finance Res. Lett.*, vol. 35, Jul. 2020, Art. no. 101599, doi: 10.1016/j.frl.2020.101599.

[41] J. B. Hollander, H. Renski, C. Foster-Karim, and A. Wiley, "Micro quality of life: Assessing health and well-being in and around public facilities in New York City," *Appl. Res. Quality Life*, vol. 15, no. 3, pp. 791–812, Jul. 2020, doi: 10.1007/s11482-019-9705-9.

[42] D. Wang, A. Al-Rubaie, B. Hirsch, and G. C. Pole, "National happiness index monitoring using Twitter for bilanguages," *Social Netw. Anal. Mining*, vol. 11, no. 1, p. 24, Dec. 2021, doi: 10.1007/s13278-021-00728-0.

[43] S. Pakhomov, N. Shah, P. Hanson, S. Balasubramaniam, S. A. Smith, and S. A. Smith, "Automatic quality of life prediction using electronic medical records," in *Proc. AMIA Annu. Symp.*, Nov. 2008, pp. 545–549.

[44] S.-K. Lee, Y.-J. Son, J. Kim, H.-G. Kim, J.-I. Lee, B.-Y. Kang, H.-S. Cho, and S. Lee, "Prediction model for health-related quality of life of elderly with chronic diseases using machine learning techniques," *Healthcare Informat. Res.*, vol. 20, no. 2, p. 125, 2014, doi: 10.4258/hir.2014.20.2.125.

[45] Z. Erdoğan and E. Namlü, "A living environment prediction model using ensemble machine learning techniques based on quality of life index," *J. Ambient Intell. Humanized Comput.*, vol. 2019, pp. 1–17, Sep. 2019, doi: 10.1007/s12652-019-01432-w.

[46] S. Khodabakhsh, N. A. Rosli, and N. A. b. Hushairi, "Relationship between wellbeing and academic achievement among university students," in *Proc. IEEE 11th Int. Conf. Eng. Educ. (ICEED)*. Kanazawa, Japan: IEEE, Nov. 2019, pp. 178–181, doi: 10.1109/ICEED47294.2019.8994824.

[47] G. Egilmez, N. Ö. Erdil, O. M. Arani, and M. Vahid, "Application of artificial neural networks to assess student happiness," *Int. J. Appl. Decis. Sci.*, vol. 12, no. 2, p. 115, 2019, doi: 10.1504/ijads.2019.098674.

[48] P. Sancho, J. M. Tomás, A. Oliver, L. Galiana, and M. Gutiérrez, "Predicting life satisfaction in Spanish adults of advancing age attending university educational programs," *J. Happiness Stud.*, vol. 21, no. 1, pp. 271–287, Jan. 2020, doi: 10.1007/s10902-019-00073-2.

[49] X. Tao, O. Chi, P. J. Delaney, L. Li, and J. Huang, "Detecting depression using an ensemble classifier based on quality of life scales," *Brain Informat.*, vol. 8, no. 1, p. 2, Dec. 2021, doi: 10.1186/s40708-021-00125-5.

[50] T. R. D. Saputri and S.-W. Lee, "Are we living in a happy country: An analysis of national happiness from machine learning perspective," in *Proc. Int. Conf. Softw. Eng. Knowl. Eng.*, Jul. 2015, pp. 174–177, doi: 10.18293/seke2015-224.

[51] E. Yaman, A. Music-Kiliç, and Z. Zerdo, "Using classification to determine whether personality profiles of countries affect various national indexes," in *Proc. Int. Conf. Control, Artif. Intell., Robot. Optim. (ICCAIRO)*, May 2018, pp. 48–52, doi: 10.1109/ICCAIRO.2018.00017.

[52] M. Kaur, M. Dhalaria, P. K. Sharma, and J. H. Park, "Supervised machine-learning predictive analytics for national quality of life scoring," *Appl. Sci.*, vol. 9, no. 8, p. 1613, Apr. 2019, doi: 10.3390/app9081613.

[53] S. Feuerriegel and J. Gordon, "News-based forecasts of macroeconomic indicators: A semantic path model for interpretable predictions," *Eur. J. Oper. Res.*, vol. 272, no. 1, pp. 162–175, Jan. 2019, doi: 10.1016/j.ejor.2018.05.068.

[54] C. Katris, "Prediction of unemployment rates with time series and machine learning techniques," *Comput. Econ.*, vol. 55, no. 2, pp. 673–706, Feb. 2020, doi: 10.1007/s10614-019-09908-9.

[55] P. Prokopowicz, D. Mikołajewski, E. Mikołajewska, and P. Kotlarz, "Fuzzy system as an assessment tool for analysis of the health-related quality of life for the people after stroke," in *Artificial Intelligence and Soft Computing* (Lecture Notes in Computer Scienc), vol. 10245, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, Eds. Cham, Switzerland: Springer, 2017, pp. 710–721, doi: 10.1007/978-3-319-59063-9_64.

[56] D. Kelly, K. Curran, and B. Caulfield, "Automatic prediction of health status using smartphone-derived behavior profiles," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 6, pp. 1750–1760, Nov. 2017, doi: 10.1109/JBHI.2017.2649602.

[57] B. Li, H. Yu, and A. Sano, "Toward end-to-end prediction of future well-being using deep sensor representation learning," in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interact. Workshops Demos (ACIIW)*, Sep. 2019, pp. 253–257, doi: 10.1109/ACIIW.2019.8925072.

[58] M. Beeksma, S. Verberne, A. van den Bosch, E. Das, I. Hendrickx, and S. Groenewoud, "Predicting life expectancy with a long short-term memory recurrent neural network using electronic medical records," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, p. 36, Dec. 2019, doi: 10.1186/s12911-019-0775-2.

[59] O. Spiga, V. Cicaloni, C. Fiorini, A. Trezza, A. Visibelli, L. Millucci, G. Bernardini, A. Bernini, B. Marzocchi, D. Braconi, F. Prischi, and A. Santucci, "Machine learning application for development of a data-driven predictive model able to investigate quality of life scores in a rare disease," *Orphanet J. Rare Diseases*, vol. 15, no. 1, p. 46, Dec. 2020, doi: 10.1186/s13023-020-1305-0.

[60] J.-C. Huang, "Using fuzzy neural network to explore the effect of internet on quality of life," in *Computational Collective Intelligence. Technologies and Applications* (Lecture Notes in Computer Science), vol. 6422, J.-S. Pan, S.-M. Chen, and N. T. Nguyen, Eds. Heidelberg, Germany: Springer, 2010, pp. 402–410, doi: 10.1007/978-3-642-16732-4_43.

[61] T. Umematsu, A. Sano, and R. W. Picard, "Daytime data and LSTM can forecast tomorrow's stress, health, and happiness," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*. Berlin, Germany: IEEE, Jul. 2019, pp. 2186–2190, doi: 10.1109/EMBC.2019.8856862.

[62] Z. Yang, D. Olszewski, C. He, G. Pintea, J. Lian, T. Chou, R. C. Chen, and B. Shtylla, "Machine learning and statistical prediction of patient quality-of-life after prostate radiation therapy," *Comput. Biol. Med.*, vol. 129, Feb. 2021, Art. no. 104127, doi: 10.1016/j.compbiomed.2020.104127.

[63] K. Kitikidou and L. Iliadis, "A multi-layer perceptron neural network to predict air quality through indicators of life quality and welfare," in *Artificial Intelligence Applications and Innovations* (IFIP Advances in Information and Communication Technology), vol. 339, H. Papadopoulos, A. S. Andreou, and M. Bramer, Eds. Heidelberg, Germany: Springer, 2010, pp. 395–402, doi: 10.1007/978-3-642-16239-8_51.

[64] U. Salma, Md. M. Hasan, and M. I. Sultana, "A regression model-based approach to identifying determining factors for GDP growth in Bangladesh," in *Proc. Int. Conf. Emerg. Trends Inf. Technol. Eng.*, Feb. 2020, pp. 1–6, doi: 10.1109/ic-ETITE47903.2020.392.

[65] J. Potter, R. Cantarero, and H. Wood, "The multi-dimensional nature of predicting quality of life," *Proc. Social Behav. Sci.*, vol. 50, pp. 781–790, Jan. 2012, doi: 10.1016/j.sbspro.2012.08.080.

[66] K. Ruggeri, E. Garcia-Garzon, Á. Maguire, S. Matz, and F. A. Huppert, "Well-being is more than happiness and life satisfaction: A multidimensional analysis of 21 countries," *Health Quality Life Outcomes*, vol. 18, no. 1, p. 192, Jun. 2020, doi: 10.1186/s12955-020-01423-y.

[67] N. Sansó, L. Galiana, A. Oliver, M. Tomás-Salvá, and G. Vidal-Blanco, "Predicting professional quality of life and life satisfaction in Spanish nurses: A cross-sectional study," *Int. J. Environ. Res. Public Health*, vol. 17, no. 12, p. 4366, Jun. 2020, doi: 10.3390/ijerph17124366.

[68] V. Capone, M. Joshanloo, and H. Scheifinger, "Predictors of life satisfaction in a large representative sample from Italy," *Current Psychol.*, vol. 40, no. 7, pp. 3609–3627, Jul. 2021, doi: 10.1007/s12144-019-00305-x.

[69] E. Dejonckheere et al., "Perceiving societal pressure to be happy is linked to poor well-being, especially in happy nations," *Sci. Rep.*, vol. 12, no. 1, Feb. 2022, Art. no. 1, doi: 10.1038/s41598-021-04262-z.

[70] M. Weckroth, S. Ala-Mantila, D. Ballas, T. Ziogas, and J. Ikonen, "Urbanity, neighbourhood characteristics and perceived quality of life (QoL): Analysis of individual and contextual determinants for perceived QoL in 3300 postal code areas in Finland," *Social Indicators Res.*, vol. 164, no. 1, pp. 139–164, Feb. 2022, doi: 10.1007/s11205-021-02835-z.

[71] C.-T. Su and F.-F. Wang, "Integrated fuzzy-connective-based aggregation network with real-valued genetic algorithm for quality of life evaluation," *Neural Comput. Appl.*, vol. 21, no. 8, pp. 2127–2135, Nov. 2012, doi: 10.1007/s00521-011-0644-0.

[72] W.-j. Hsieh and B. Chen-heng Yang, "The pursuit of happiness: An empirical study of BLI of OECD countries," in *Proc. Int. Conf. Orange Technol. (ICOT)*, Dec. 2015, pp. 83–87, doi: 10.1109/ICOT.2015.7498481.

[73] L. Carlsen, "Happiness as a sustainability factor. The world happiness index: A posetic-based data analysis," *Sustainability Sci.*, vol. 13, no. 2, pp. 549–571, Mar. 2018, doi: 10.1007/s11625-017-0482-9.

[74] H. I. Okagbue, P. E. Oguntunde, S. A. Bishop, P. I. Adamu, E. M. Akhmetshin, and C. O. Iroham, "Significant predictors of henley passport index," *J. Int. Migration Integr.*, vol. 22, no. 1, pp. 21–32, Mar. 2021, doi: 10.1007/s12134-019-00726-4.

[75] J. Yoon, "Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach," *Comput. Econ.*, vol. 57, no. 1, pp. 247–265, Jan. 2021, doi: 10.1007/s10614-020-10054-w.

[76] A. Pawliczek, J. Kurowska-Pysz, and R. Smilnak, "Relation between globe latitude and the quality of life: Insights for public policy management," *Sustainability*, vol. 14, no. 3, p. 1461, Jan. 2022, doi: 10.3390/su14031461.

[77] S. Hajduk, "Using multivariate statistical methods to assess the urban smartness on the example of selected European cities," *PLoS ONE*, vol. 15, no. 12, Dec. 2020, Art. no. e0240260, doi: 10.1371/journal.pone.0240260.

[78] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Sci. Amer.*, vol. 284, no. 5, pp. 34–43, 2001.

[79] M. Taye, "Understanding semantic web and ontologies: Theory and applications," *J. Comput.*, vol. 2, Jun. 2010.

[80] T. Rahman, R. C. Mittelhammer, and P. R. Wandschneider, "Measuring quality of life across countries: A multiple indicators and multiple causes approach," *J. Socio-Economics*, vol. 40, no. 1, pp. 43–52, Feb. 2011, doi: 10.1016/j.socec.2010.06.002.

[81] L. Ghahremanloo, J. A. Thom, and L. Magee, "An ontology derived from heterogeneous sustainability indicator set documents," in *Proc. 17th Australas. Document Comput. Symp.* Dunedin, New Zealand: ACM Press, Dec. 2012, pp. 72–79, doi: 10.1145/2407085.2407095.

[82] S. Nalchigar and M. S. Fox, "An ontology for open 311 data," in *Proc. AAAI Workshop*, 2014, pp. 29–35.

[83] J. Hunter, "An ontological framework to support evidence-based policy making using global city indicators," in *Proc. IEEE 1st Int. Smart Cities Conf. (ISC2)*, Oct. 2015, pp. 1–7, doi: 10.1109/ISC2.2015.7366209.

[84] M. S. Fox, "The role of ontologies in publishing and analyzing city indicators," *Comput., Environ. Urban Syst.*, vol. 54, pp. 266–279, Nov. 2015, doi: 10.1016/j.compenvurbsys.2015.09.009.

[85] P. Lula and M. Hamerska, "Ontology-based models for quality of life measurement and comparison," *Int. J. Quality Res.*, vol. 10, no. 3, pp. 595–606, 2016, doi: 10.18421/IJQR10.03-11.

[86] M. S. Fox and C. J. Pettit, "On the completeness of open city data for measuring city indicators," in *Proc. IEEE 1st Int. Smart Cities Conf. (ISC2)*, Oct. 2015, pp. 1–6, doi: 10.1109/ISC2.2015.7366147.

[87] F. Benvenuti, C. Diamantini, D. Potena, and E. Storti, "An ontology-based framework to support performance monitoring in public transport systems," *Transp. Res. C, Emerg. Technol.*, vol. 81, pp. 188–208, Aug. 2017, doi: 10.1016/j.trc.2017.06.001.

[88] T. S. J. Smith and L. Reid, "Which 'being' in wellbeing? Ontology, wellness and the geographies of happiness," *Prog. Hum. Geography*, vol. 42, no. 6, pp. 807–829, Dec. 2018, doi: 10.1177/0309132517717100.

[89] M. S. Fox, "The semantics of populations: A city indicator perspective," *J. Web Semantics*, vol. 48, pp. 48–65, Jan. 2018, doi: 10.1016/j.websem.2018.01.001.

[90] M. Katsumi and M. Fox, "An ontology-based standard for transportation planning," in *Proc. CEUR Workshop*, 2019.

[91] S.-K. Hu and G.-H. Tzeng, "Strategizing for better life development using OECD well-being indicators in a hybrid fuzzy MCDM model," *Int. J. Fuzzy Syst.*, vol. 19, no. 6, pp. 1683–1702, Dec. 2017, doi: 10.1007/s40815-017-0328-8.

[92] V. Edgar, C. La Place, J. Schmidt, A. Bansal, and S. Bansal, "SustainOnt: An ontology for defining an index of neighborhood sustainability across domains," in *Proc. Int. Workshop Semantic Big Data*, Jun. 2020, pp. 1–13, doi: 10.1145/3391274.3393640.

[93] M. S. Fox, "An ontology engineering approach to measuring city education system performance," *Expert Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115734, doi: 10.1016/j.eswa.2021.115734.

[94] V. Estrada-Galinanes and K. Wac, "Visions and challenges in managing and preserving data to measure quality of life," in *Proc. IEEE 3rd Int. Workshops Found. Appl. Self Syst.*, Sep. 2018, pp. 92–99, doi: 10.1109/FAS-W.2018.00031.
[95] J. W. Knopf, "Doing a literature review," *PS, Political Sci. Politics*, vol. 39, no. 1, pp. 127–132, Jan. 2006, doi: 10.1017/s1049096506060264.
[96] J. Rowley and F. Slack, "Conducting a literature review," *Manage. Res. News*, vol. 27, no. 6, pp. 31–39, Jun. 2004, doi: 10.1108/01409170410784185.
[97] E. T. Rother, "Revisão sistemática X revisão narrativa," *Acta Paulista Enfermagem*, vol. 20, no. 2, pp. 5–6, Jun. 2007, doi: 10.1590/s0103-21002007000200001.
[98] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele Univ., Keele, U.K., Durham Univ., Durham, U.K., Tech. Rep. EBSE-2007-01, 2007.
[99] S. O. Rutstein, "The DHS wealth index: Approaches for rural and urban areas," Macro Int. Inc., Calverton, MD, USA, DHS Work. Papers 60, 2008.

AYOUB JANNANI received the Bachelor of Science and Techniques degree in computer science from the Faculty of Sciences and Techniques, Sultan Molay Slimane University, Beni Mellal, Morocco, in 2018, and the Master of Science degree in data science and big data from the Faculty of Sciences Ben M'Sick, Hassan II University of Casablanca, Morocco, in 2020. He is currently pursuing the Ph.D. degree in computer science.

His research endeavors are concentrated on leveraging artificial intelligence to enhance quality of life and wellbeing, specifically through the use of natural language processing, machine learning, deep learning, and time series analysis.

NAWAL SAEL received the engineering degree in software engineering from ENSIAS, Morocco, in 2002.

Since 2012, she has been a Teacher-Researcher, achieving the rank of an Authorized Professor, in 2014, and has been a Professor of higher education with the Department of Mathematics and Computer Science, Faculty of Sciences Ben M'Sick, Casablanca, since 2020. Her research interests include data mining, educational data mining, machine learning, deep learning, and the Internet of Things.

FAOUZIA BENABBOU was a Teacher-Researcher, in 1994. She ascended to the position of an Authorized Professor, in 2008, and has been appointed as a Professor of higher education with the Department of Mathematics and Computer Science, Faculty of Sciences Ben M'Sick, Casablanca, since 2015. She plays an integral role in the Information Technology and Modeling Laboratory, where she leads the Cloud Computing, Network, and Systems Engineering (ICCNSE) Team. Her research endeavors are concentrated in the fields of cloud computing, data mining, machine learning, and natural language processing.