

Finze, Nikola; Jechle, Deinera; Faußer, Stefan; Gewalt, Heiko

Article — Published Version

How are We Doing Today? Using Natural Speech Analysis to Assess Older Adults' Subjective Well-Being

Business & Information Systems Engineering

Suggested Citation: Finze, Nikola; Jechle, Deinera; Faußer, Stefan; Gewalt, Heiko (2024) : How are We Doing Today? Using Natural Speech Analysis to Assess Older Adults' Subjective Well-Being, Business & Information Systems Engineering, ISSN 1867-0202, Springer Fachmedien Wiesbaden, Wiesbaden, Vol. 66, Iss. 3, pp. 321-334,
<https://doi.org/10.1007/s12599-024-00877-4>

This Version is available at:

<https://hdl.handle.net/10419/315765>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>



How are We Doing Today? Using Natural Speech Analysis to Assess Older Adults' Subjective Well-Being

Nikola Finze · Deiner Jechle · Stefan Faußer · Heiko Gewalt

Received: 23 June 2023 / Accepted: 17 April 2024 / Published online: 8 June 2024
© The Author(s) 2024

Abstract The research presents the development and test of a machine learning (ML) model to assess the subjective well-being of older adults based solely on natural speech. The use of such technologies can have a positive impact on healthcare delivery: the proposed ML model is patient-centric and securely uses user-generated data to provide sustainable value not only in the healthcare context but also to address the global challenge of demographic change, especially with respect to healthy aging. The developed model unobtrusively analyzes the vocal characteristics of older adults by utilizing natural language processing but without using speech recognition capabilities and adhering to the highest privacy standards. It is based on theories of subjective well-being, acoustic phonetics, and prosodic theories. The ML models were trained with voice data from volunteer participants and calibrated through the World Health Organization Quality of Life Questionnaire (WHOQOL), a widely accepted tool for assessing the subjective well-being of human beings. Using WHOQOL scores as a proxy, the developed model provides accurate numerical estimates of individuals' subjective well-being.

Different models were tested and compared. The regression model proves beneficial for detecting unexpected shifts in subjective well-being, whereas the support vector regression model performed best and achieved a mean absolute error of 10.90 with a standard deviation of 2.17. The results enhance the understanding of the subconscious information conveyed through natural speech. This offers multiple applications in healthcare and aging, as well as new ways to collect, analyze, and interpret self-reported user data. Practitioners can use these insights to develop a wealth of innovative products and services to help seniors maintain their independence longer, and physicians can gain much greater insight into changes in their patients' subjective well-being.

Keywords Artificial intelligence · Machine learning · Natural language processing · Older adult · Subjective well-being assessment · WHOQOL-OLD, WHOQOL-BREF · Voice analysis

1 Introduction

Taking care of the aging generation is a societal task that puts nations around the world under pressure. Older adults (as per the WHO defined as people 60 + years of age) who profusely suffer from chronic diseases and declining health in general need care and attention, which collides with the trend towards smaller and geographically dispersed families, specifically in developed economies. The resulting situation is alarming: in Germany, around one-third of all persons over 65 live alone, and this level is expected not to decrease (Statistisches Bundesamt 2023). Simultaneously, the number of professional caretakers is continuously declining (Flake et al. 2018). This is a worldwide

Accepted after three revisions by the editors of the Special Issue.

N. Finze (✉) · D. Jechle · S. Faußer · H. Gewalt
Center for Research on Service Sciences (CROSS), Faculty of
Information Management, Neu-Ulm University of Applied
Sciences, Wileyst. 1, 89231 Neu-Ulm, Germany
e-mail: nikola.finze@hnu.de

D. Jechle
e-mail: deiner.jechle@hnu.de

S. Faußer
e-mail: stefan.fausser@hnu.de

H. Gewalt
e-mail: heiko.gewald@hnu.de

phenomenon that is not restricted to Germany. As the global population ages, the subjective well-being of older adults has gained substantial attention from healthcare providers, policymakers, and society (United Nations 2019; National Institute on Aging 2021). This global challenge is underscored by the United Nations (UN) commitment to the *UN Decade of Healthy Aging* objectives, a global initiative to promote the health and quality of life of older individuals worldwide, inaugurated in 2021 (World Health Organization 2023). An individual's subjective well-being is a complex multidimensional construct encompassing physical and mental health, social support, and life satisfaction (Diener 1984; Cooke et al. 2016).

For most seniors, a key well-being factor is to live a self-sustained life in their familiar environment for as long as possible (Dierx 2019). A challenge to this demand is the general decline of physical and mental abilities as a natural consequence of aging (Gaertner et al. 2023). Negative developments might go unnoticed if the traditional weekly visit or phone call – including the standard question “How are you doing today?” – is the only external measure of an individual's well-being. Indeed, assessing a person's ‘real’ well-being is difficult as this construct is entirely subjective. To address this issue, psychologists have developed and tested numerous measurements based on structured questionnaires (Cooke et al. 2016). The World Health Organization (WHO) made a great effort to develop an internationally validated instrument to measure individuals' perceived quality of life, the WHOQOL (The WHOQOL Group 1998b). It was a joint development of 15 international research centers, tested with more than 4,500 participants, and translated into more than 70 languages. It is generally considered a valid instrument producing adequate results (The WHOQOL Group 1998a; Cooke et al. 2016). The WHO also developed a specialized instrument to assess the perceptions of senior citizens, the WHOQOL-OLD (Power et al. 2005), which supplements the general WHOQOL.

However, although the WHOQOL instruments provide a comprehensive and dependable assessment of an individual's subjective well-being, they are not suitable for frequent use due to the large number of questions (100) the participants need to answer. Thus, if this assessment could be done automatically, unobtrusively, and continuously, relatives and professional caregivers would be able to react more timely to prevent adverse outcomes in case of declines in subjective well-being (Artola et al. 2021). Additionally, professional caregivers are becoming scarce (Ribeiro et al. 2021; Flake et al. 2018), and the optimal deployment of these skilled workers is becoming increasingly important for national healthcare systems. The increasing number of older people in combination with the decrease in caregivers calls for technological solutions to

ease the negative impacts of this increasing imbalance (Martinho et al. 2019; Czaja and Ceruso 2022).

One way to address the aforementioned challenges could be the continuous automated assessment of seniors' subjective well-being through natural speech analysis. Technological advancements in machine learning (ML) and natural language processing (NLP) have demonstrated promising potential in deriving subconscious information from the human voice (Zunic et al. 2020). In the healthcare context, for example, NLP has been used to analyze patient feedback, gauge sentiment, and even detect early signs of diseases from patients' speech patterns and use of language (DeSouza et al. 2021; Khanbhai et al. 2021; Beltrami et al. 2018; Perez et al. 2018). Furthermore, within linguistic research, acoustic phonetics theory, which focuses on the sounds of speech (Ververidis and Kotropoulos 2006), and prosody theory, which focuses on elements such as pitch, duration, and intensity of speech (Hubbard et al. 2017; Barnes and Shattuck-Hufnagel 2022; Ladd 2008), are critical for the expression of emotions (Corrales-Astorgano et al. 2019), and the identification of an individual's emotional state (Bhavan et al. 2019; Rusz et al. 2011; Godino-Llorente and Gomez-Vilda 2004).

However, this research stream has primarily focused on specific aspects of emotions, such as identifying depression (Rejaibi et al. 2022; Lin et al. 2020; An et al. 2019), early signs of diseases from patients' speech patterns (DeSouza et al. 2021; Khanbhai et al. 2021; Beltrami et al. 2018; Perez et al. 2018), or challenges in emotion recognition among older adults (Schuller et al. 2020), without fully harnessing the potential of those technologies for the automated assessment and monitoring of subjective well-being. By utilizing research on NLP analytic capabilities, we aim to address the global challenge of healthy aging by assessing subjective well-being through speech analysis. Therefore, we formulate the following research question: *How can older adult's subjective well-being be assessed from natural speech?*

Our work contributes to research on advances in the automatic assessment of subjective well-being through ML. We present an innovative, patient-centered model that uses NLP to enable data-driven care and support for the aging generation. The model is a novel way to assess a human's subjective well-being unobtrusively, continuously, and automatically and could provide family and professional caretakers with timely and comprehensive overviews of a senior's subjective well-being. It could trigger interventions as needed, thus supporting the optimal utilization of the scarce resource of professional caretakers. It would also provide ease-of-mind for family members if frequent interaction with the senior is not possible, for instance, due to geographic distance.

The ML model developed to answer the research question is theoretically rooted in subjective well-being theory, as well as theories of prosody and acoustic phonetics. This foundation is combined with NLP techniques, including feature extraction from speech and regression models. Workshops with stakeholders have been conducted to derive requirements for the technical solution and ensure the usefulness of the practical contribution. The training dataset for the model was created by recording the voices of German-speaking senior citizens while they completed the WHOQOL questionnaires. The ML training of the participants' voices was calibrated with the results of their WHOQOL assessment.

In the following, this paper elaborates on the theoretical basis of this research, provides an overview of the current state of technological development, and describes the data collection procedures as well as the methodological approach of the ML algorithm. Based on the evaluation of the results, the performance of the ML model is discussed, and the implications for theory and practice are presented. The paper closes by explicating its limitations, providing avenues for future research, and its conclusion.

2 Theoretical Background and Related Work

Our research is positioned at the intersection of subjective well-being measurement and the computational analysis of emotion recognition from natural speech. We address our research question based on three theoretical foundations: subjective well-being, prosody, and acoustic phonetics. The integration of these theories poses that the perceived human well-being is closely connected to a person's feelings and emotions. These are subconsciously transmitted through the human voice, which allows NLP methods to detect and quantify them (Lin et al. 2020). This section discusses the theoretical foundations and provides an overview of related research.

2.1 Subjective Well-being

Historically, well-being was defined as the absence of disease and disability (Cooke et al. 2016). However, in 1948, the WHO stated, "*health is a state of complete physical, mental, and social well-being and not merely the absence of disease and infirmity*" (World Health Organization 2020). Since then, research on subjective well-being has received considerable attention, particularly in psychology, and is now moving toward a multidimensional concept that encompasses an individual's optimal functioning and satisfaction in various aspects of life (The WHOQOL Group 1998b). Among the different approaches to assess subjective well-being, hedonic approaches

emphasize pleasure and happiness, with subjective well-being as a prominent model consisting of life satisfaction, absence of negative affect, and presence of positive affect (Cooke et al. 2016). Furthermore, eudaimonic approaches to well-being propose that psychological health is achieved "*by fulfilling one's potential, functioning at an optimal level*" (Cooke et al. 2016), which includes living in a way that fulfills one's potential and leads to personal growth and development (Lent 2004).

While in literature, the terms quality of life and subjective well-being are often used interchangeably, quality of life researchers conceptualize subjective well-being more broadly to include hedonic (presence of pleasure, absence of pain) and eudaimonic perspectives (living a meaningful and purposeful life) and are based on physical, psychological, and social aspects (Vik and Carlquist 2018; Cooke et al. 2016). Following these arguments, we broadly conceptualized subjective well-being for this study to include physical, mental, and social dimensions (Diener et al. 2009; Cooke et al. 2016).

As subjective well-being strongly influences physical and mental health, resilience, and many more essential aspects of human life (Ryff 2014; Cooke et al. 2016; Yıldırım and Çelik Tanrıverdi 2020), the WHO put great effort into developing a standardized measure for it (Huppert and So 2013; Keyes 2005). In 1998, the WHO presented the quality of life instruments (WHOQOL) developed in an international study program, including focus groups with patients and medical experts worldwide (World Health Organization 2012). The WHOQOL has the advantage that it was developed cross-culturally and, therefore, can be used in international studies without additional adaptations. This opens avenues for further research, which can enhance the generalizability of the findings to other languages and cultural contexts (Cooke et al. 2016). It is widely regarded as a high-quality patient-centered tool, successfully assessing individuals' subjective well-being in different research areas (Skevington and McCrate 2012).

The instrument comprises four domains: physical, psychological, social relationships, and environment. The physical domain focuses on pain and energy, the psychological domain on self-esteem, the social relationships on personal relationships, and the environmental domain on safety and financial resources. In this way, the WHOQOL instrument recognizes that well-being is influenced by multiple interconnected factors and shaped by the collective impact of various dimensions (The WHOQOL Group 1998b; World Health Organization 2012).

The WHOQOL is supplemented by a set of specific instruments like the WHOQOL-HIV, which is calibrated for people infected with HIV, or the WHOQOL-OLD, which is context-specific to the subjective well-being of people over 60 years of age (Centers for Disease Control

Prevention 2012; World Health Organization 2012). The WHOQOL-OLD reflects the unique situation and challenges the older population faces (Diener 1984). The questionnaire comprises 24 questions across six dimensions: sensory abilities, autonomy, past, present, and future, social participation, death, and intimacy (Power et al. 2005). It is considered a robust and reliable instrument used in various countries and contexts, making it an essential tool in gerontological research (Lucas-Carrasco 2012; Chachamovich et al. 2008; Conrad et al. 2014).

The WHOQOL is comprised of 100 questions, necessitating considerable time for thoughtful and accurate responses. Consequently, an abbreviated version known as the WHOQOL-BREF was developed to address this issue. It is reduced to 26 questions and generates results with reliability comparable to the longer WHOQOL (The WHOQOL Group 1998a). To understand older adults' well-being, both instruments (WHOQOL-BREF and WHOQOL-OLD) must be combined (i.e., the participants must answer both questionnaires). To enhance readability, we abbreviate the combined use of both instruments as 'QOLs' in the remainder of this paper.

2.2 Transmission of Human Feelings and Emotions via Natural Speech

The human voice expresses both conscious and subconscious information like feelings and emotions (Li et al. 2019). Therefore, the ability to analyze the human voice to derive the speaker's subconscious feelings can be a solution to assess an individual's subjective well-being in an unintrusive way through an automated ML algorithm. Acoustic phonetics and prosody provide the foundation for extracting the relevant features from speech data via NLP methods.

2.2.1 Acoustic Phonetic Theory

Acoustic phonetics is the study of the physical properties of speech sounds, including their production, transmission, and perception, apart from the actual words spoken (Ladefoged and Johnson 2014). It focuses on how speech sounds are formed as air is forced out of the lungs and examines phonemes' articulatory and auditory characteristics, focusing on formants, pitch, intensity, spectral characteristics, and duration (Ververidis and Kotropoulos 2006). According to previous research, variations in these features indicate changes in a person's emotional state, health condition, or well-being (Godino-Llorente and Gomez-Vilda 2004; Ruzs et al. 2011). For instance, stress or emotional distress often manifests as changes in the pitch, intensity, or rhythm of speech (Reiner 2013). Research has also demonstrated the role of acoustic

phonetics in identifying emotional states such as sadness, anger, happiness, and fear (Bhavan et al. 2019; Tariq et al. 2019).

2.2.2 Prosodic Theory

The prosodic theory focuses on central aspects of speech, including stress, intonation, rhythm, and phrasing (Ladd 2008; Barnes and Shattuck-Hufnagel 2022). It explores how pitch, duration, and intensity variations shape speech's melody, rhythm, and emphasis, conveying linguistic and affective information (Hubbard et al. 2017). Prosody is crucial in conveying meaning, expressing emotions, and structuring discourse (Corrales-Astorgano et al. 2019). For instance, a change in intonation can turn a statement into a question. By studying prosodic features, researchers gain insights into speech's expressive and pragmatic aspects (Weed and Fusaroli 2020).

Given the intricate nature of speech and its emotional content, approaches that incorporate both acoustic phonetic and prosodic features offer a more comprehensive and accurate analysis. This multimodal approach, which combines the physical properties of speech sounds (acoustic features) and the rhythm, stress, and intonation of speech (prosodic features), is considered more effective and closer to the natural complexity of human speech (Byun et al. 2021). A multimodal approach allows for a nuanced analysis of speech, capturing both the linguistic and affective information conveyed in speech, thereby providing a more realistic representation of human emotional expression (Schuller et al. 2020; An et al. 2019).

2.3 Applying NLP to Assess Feelings and Emotions from Natural Speech

The use of NLP to detect human well-being is a comparatively new area of research. This new area presents significant interdisciplinary challenges, combining the complex theories of emotion expression in human speech (i.e., acoustic phonetics and prosody) with advanced computational techniques. While the literature has noted conversational agents and chatbots (e.g., Ahmad et al. 2022) for assessing subjective well-being, human voice analysis offers a particularly intriguing avenue for research. Published research focuses on detecting feelings, emotions, and mental health issues through voice analysis, using various approaches to capture the complex signals conveyed by the speaker's voice. While these studies differ in their specific research goals, they collectively contribute to a broader understanding of how subjective well-being can be measured non-invasively through natural speech. Three main research streams have emerged: depression detection,

speech emotion recognition, and subjective well-being measurement from speech.

2.3.1 Depression Detection

NLP techniques are used in depression detection due to the measurable differences in speech signals of depressed patients, as well as the non-intrusive, remote, and low-cost nature of speech detection (Rejaibi et al. 2022; Lin et al. 2020). Early studies have focused on elements such as prosody and spectral analysis, typically paired with conventional ML techniques (Wu et al. 2023). Sanchez et al. (2011) used prosodic and spectral speech features to identify differences between healthy and depressed individuals using a support vector machine (SVM), yielding a prediction accuracy of over 81%. Even with advanced techniques, SVMs remained effective in modeling and classifying acoustic signals that indicate depression in real-time with an accuracy of 90% (Yalamanchili et al. 2020). However, more complex ML algorithms, specifically deep learning, are now used to capture more nuanced features in speech (Wu et al. 2023). Lin et al. (2020) integrated a bidirectional LSTM and a one-dimensional convolutional neural network (1D-CNN) to capture temporal and spectral features from speech data, leading to greater accuracy in detecting depression. Their proposed model achieved up to 84% accuracy by employing purely acoustic features. In a similar vein, long short-term memory (LSTM) models that utilized mel-frequency cepstrum coefficient (MFCC) features were able to recognize patterns indicative of different levels of depression over time with approximately 76% accuracy (Rejaibi et al. 2022).

2.3.2 Emotion Detection

Emotion detection from speech, also referred to as speech emotion recognition (SER), focuses on representing emotions using numerical value feature sets extracted from speech data (Pentari et al. 2024). ML algorithms, like SVMs, Decision Trees, or Random Forest, have been used to detect emotional states from natural speech based on analysis of the vocal features (Suresh et al. 2023). While feature sets yield good outcomes, nuances in speech often lack sufficient representation in such sets (Wang et al. 2015; Pentari et al. 2024). Recent advances in SER have shown the power of deep learning algorithms for detecting emotions from speech. Similarly, Tariq et al. (2019) developed a 2D-CNN to detect seven basic emotions (calm, happy, sad, angry, fearful, disgust, and surprise) in the speech of elderly patients, with an overall accuracy of up to 95%. In related research, multimodal approaches using cascaded LSTM recurrent neural networks (RNN) were used by Gupta et al. (2022) to detect stressed and

unstressed states of test subjects with an accuracy of 91%. Schuller et al. (2020) demonstrated the effectiveness of employing acoustic and prosodic features in a CNN and LSTM RNN approach to detect arousal and valence from older adults' speech samples with 72% accuracy.

Research on emotion detection from speech is still comparatively young and the proposed feature sets using statistical and structural information extracted from speech signals to detect emotional states continuously grows (Pentari et al. 2024).

2.3.3 Measurement of Subjective Well-Being

Traditionally, questionnaires are used to assess a person's subjective well-being. Advances in ML/NLP have led to increased research activity on the automatic detection of subjective well-being from speech. However, the number of published research papers combining complex speech-emotion theories with computational techniques is comparatively small.

Kim et al. (2019) report on an NLP method to assess human subjective well-being from natural language by measuring three distinct constructs: anxiety, mood, and sleep quality. Their algorithm uses a 41-dimensional supervector of various features such as MFCCs, perceptual linear prediction (PLP), prosody, and voice quality-related features. It shows a predictive capacity of 41% for anxiety, 44% for sleep quality, and 38% for mood. Nakagawa et al. (2020) proposed a promising approach based on a 3D-CNN to estimate the quality of life of a speaker with an accuracy of 71%. Unfortunately, no more profound elaboration on the measurement items is provided in the article, and no peer-reviewed follow-up work could be located.

3 Requirements for the Technical Solution

Paying attention to the subjective well-being of older people, especially those who live alone, is essential to enable them to live independently longer. While it is natural for everyone to have good days and bad days, close relatives and designated caregivers of senior citizens need to notice a steady or sudden decline in subjective well-being to provide the necessary assistance. Traditionally, the well-being of another person is assessed through face-to-face conversations. In today's societies, however, people generally meet with each other less than they would like. Technological solutions can alert family members and caregivers when a problem arises and trigger an intervention.

To provide a sustainable and practical impact solution, it must meet several requirements derived through workshops with seniors, professional caregivers, and general

practitioners who treat elderly patients. These requirements workshops were conducted in a free discussion moderated by the project team, with no restrictions or boundaries on the final solution. After collecting these basic requirements, different technological options were discussed to fulfill the stated demands. It was concluded that an ML model based on voice analysis would be the most promising approach—accordingly, some corresponding requirements needed to be formulated for this specific technology by the project team. Table 1 provides an overview of the requirements based on the different perspectives. These requirements guided the development of this research’s ML model. A more profound elaboration on concrete design principles and the system engineering part of the research is not the subject of this publication.

4 Study Design and Data Collection

No data set to address the research question is currently publicly available. Therefore, a new data set had to be created. Senior citizens were asked to participate in a data collection exercise to collect the necessary training data. The tasks were to complete the two QOL questionnaires (WHOQOL-BREF and WHOQOL-OLD) and participate in a recorded interview. The inclusion criteria were: (1) over 60 years of age; (2) being able to live independently; (3) no mental/cognitive challenges.

Before starting the interviews, the study was explained, questions were answered, and the participant signed the data privacy agreement. The interviews took between 20 and 86 min ($\bar{O} = 42$). The interviewers asked some warm-up questions about the interviewees’ lives in general and their current subjective well-being, followed by the predefined questions of the QOLs. The audio recordings were cut and labeled with the corresponding QOL scores calculated from the respective QOL questionnaires.

Data collection was conducted from January to June 2023. Volunteers were recruited from local senior circles, institutions where seniors meet weekly to undertake charity projects, talk to each other, listen to lectures, and play games. 32 interviewees, comprising 20 females and 12 males between 60 and 88, participated in this study. According to the applicable ethics commission’s regulations, no approval was necessary as the risk assessment of this study indicated no risk or downside potential for the participants.

5 Model Development

This research discusses the development of an ML model to detect older adults’ subjective well-being -calibrated to

the QOLs- from natural language. The model analyzes an audio file of free speech from an individual and predicts the corresponding perceived well-being on a scale from 0 to 100 as per the QOLs. This section provides details regarding data labeling and feature extraction.

5.1 Data Labeling

The interviewees’ responses to the QOLs were used to label the data. Each participant answered the predefined questions across the six domains. The answers within each domain were scored and aggregated into an overall domain value between 0 and 100, according to the WHOQOL manual, whereas higher scores indicate better well-being (Conrad et al. 2016). In the next step, the domain results were aggregated into one overarching score representing the individuals’ general subjective well-being. This calculated subjective well-being score was used to label each participant’s voice recording. This approach enables a methodologically sound evaluation of the individual’s subjective well-being and allows for corresponding comparisons.

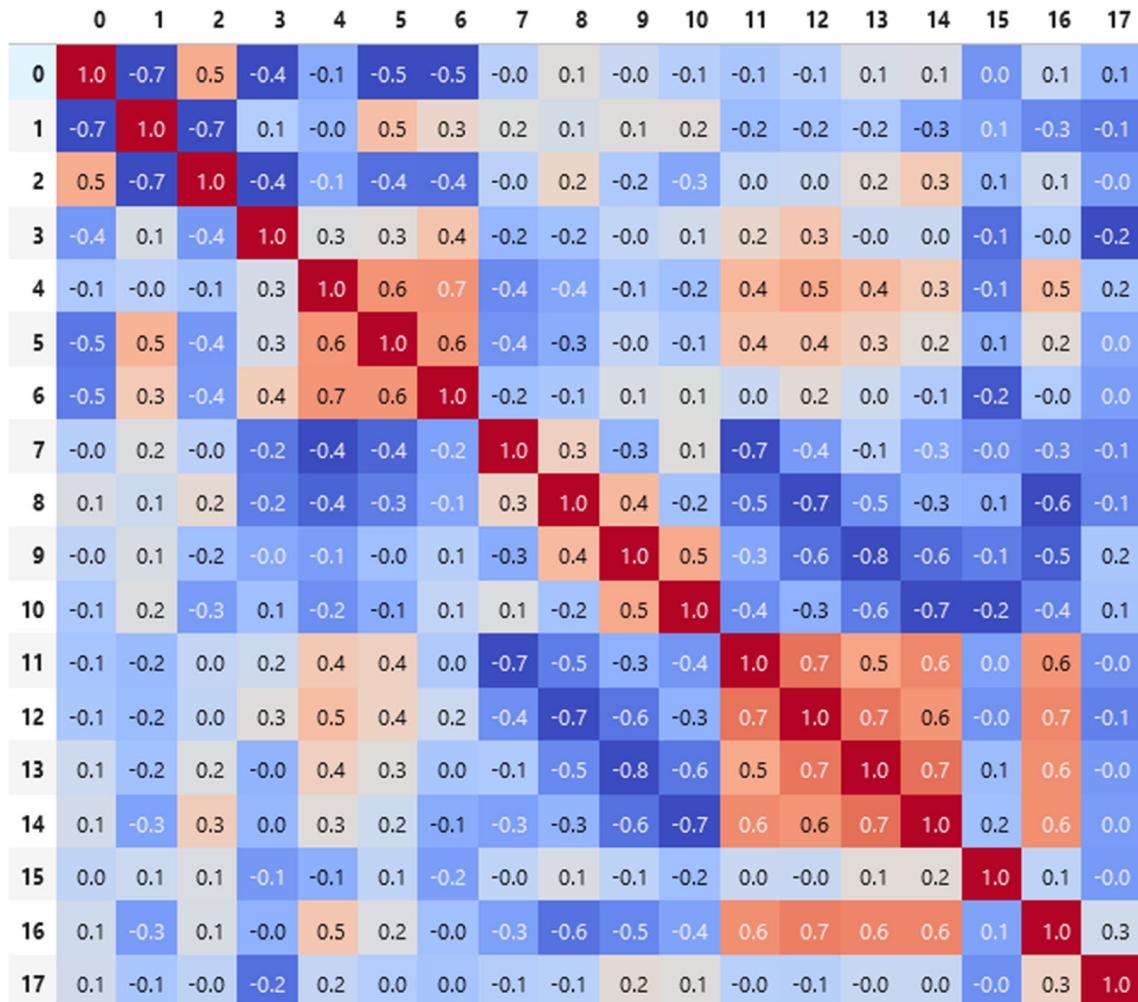
5.2 Feature Extraction

The average length of the free-speech audio files used for analysis was 30 s, with a sampling rate of 44.100 Hz. Overall, 18 acoustic phonetic and prosodic features that theoretically should indicate the speaker’s subjective well-being were extracted by employing the librosa library (McFee et al. 2015). This includes one temporal feature, the differences in speech pauses, and multiple spectral features (e.g., MFCCs). Prosodic features, such as the speech pauses, capture speech’s rhythm, stress, and intonation. These features provide insights into the temporal dynamics of speech and can be used to detect emotional states in speech (Khodabakhsh et al. 2015; Rathina et al. 2012; Sanchez et al. 2011).

The feature correlation plot of the 18 extracted features can be seen in Fig. 1. Those are: the mean of four MFCCs (0–3 in the correlation plot), the mean and standard deviation of the spectral roll-off (Klapuri and Davy 2006) for estimating the minimum spectral mass (4 and 5), the mean of the spectral centroid for estimating the mean of the spectral mass (6), the mean and standard deviation of four chromagrams, corresponding to four pitch classes (7–10; 11–14), the number of speech pauses (McFee et al. 2015) larger than 1.5 times of the median of it (15) and the first and third quantiles of the fundamental frequencies (de Cheveigne and Kawahara 2002) of the speakers (16–17). A complete description of the features can be found in the Appendix (available online via <http://link.springer.com>). Analyzing the correlations, some of the extracted features are moderately correlated but within the feature classes

Table 1 Overview of requirements

User experience and interface	Monitoring and caregiver usability	Data protection and reliability	Technology-specific requirements
Effortless user experience for seniors	Real-time or highly frequent monitoring of seniors' subjective well-being	Compliance with data protection regulations	Functionality with limited audio quality
No additional training required	Prompt/timely alerts in critical situations	Principle of voluntary use	Maintenance of accuracy and cost-effectiveness
Consistent user interface independent of updates	Time-series function for historical analysis	User anonymity preservation	Implementation and maintenance feasibility
Non-intrusive design	Easy usability for caregivers	Consistent delivery of accurate and reliable results	
	Seamless communication between caregivers and seniors		

**Fig. 1** Feature correlation plot

only, e.g., the standard deviations of the chromagrams (11–14). None of the extracted features are strongly correlated. Hence, all features add information, none are duplicates.

The continuous speech signal was divided into discrete frames of equal length based on the principle of short-time spectral analysis. Since speech is considered static for 5 to

25 ms, a window shift of 23 ms was used, employing a hop size of 1024 (Logan 2000).

5.3 Model Training

Regression models for forecasting the subjective well-being from speech data were evaluated: Four classes of

regression models were considered for the quantitative estimations of the QOL score. The evaluated regression models are: *Support vector regression* (SVR) (Drucker et al. 1996), *Lasso regression* (LR) (Tibshirani 1996, 2011), *random forest regression* (RFR) (Breiman 2001) and *XGBoost regression* (XGBR) (Chen and Guestrin 2016). Two of those (RFR and XGBR) are tree-based ensemble methods that often achieve high performances in machine learning competitions, such as those offered via Kaggle. However, when the data is small enough, the support vector machines, including SVR, perform better than the tree-based methods (Drucker et al. 1996; Boser et al. 1992). Lastly, LR is a variant of linear regression that contains feature selection. Also, SVR and XGBR implicitly make feature selections but allow for nonlinear relations between the target, the subjective well-being score, and input variables.

The models were trained and tested using a group-stratified shuffle split cross-validation with 60% for the training data, 40% for the test data, and 100 splits, where the groups are the participants. This ensures that the models were trained and tested with different, non-overlapping participants.

The hyperparameters of the models were found with a randomized grid search. Because the four classes of regression models evaluated mostly have a different set of hyperparameters, we will not list extensive tables of the evaluations in this paper but instead focus on describing which hyperparameters are the best models of each class (see Sect. 6).

5.4 Bias

A small number of recordings may cause a bias in the machine learning models, such as assigning identical subjective well-being scores for persons of the same gender. To avoid this, we tested our data to determine whether the collected demographic factors (gender, age, relationship status, and education level) affect the proposed subjective well-being. For all these factors, we found no statistically significant dependency.

6 Results

The performance of the regression models was evaluated using the metrics mean absolute error (MAE) and root mean squared error (RMSE). While both metrics can be roughly described as an average deviation from the real subjective well-being scores, the RMSE is sensitive to outliers because of the square in the term, while the MAE is less sensitive to outliers (Chicco et al. 2021). The results of the regression models are provided in Table 2.

All the evaluated models perform better than random ('educated') guesses. Those guesses were randomly distributed according to the subjective well-being score provided by the data. The best-performing model, support vector regression (SVR), has an increase of about 86%, and the second-best, the random forest regression (RFR), has an increase of about 52% in MAE to the random guesses. The best SVR model employed a radial basis function (RBF) kernel with a scaled gamma and a regularization value (typically called C) of 0.5 as hyperparameters, while the best RFR model had a maximum tree depth of six and a ten samples minimum tree split. Out of the trained models, the Lasso regression performed worst but still better than the random guesses. From this, it can be inferred that the subjective well-being score is not linearly dependent on the extracted features. This aligns with the finding that SVR and RFR, both able to capture non-linearities, had higher MAEs.

During the training of the models, the feature importances were collected utilizing the feature permutation importances technique (Breiman 2001). The feature importances for the SVR model can be seen in Fig. 2.

In the plot, 14 of the 18 extracted features are visible and ranked according to a change in score (here: MAE) when the features are randomly permuted. The remaining four features are omitted because their score change was close to zero. The feature importances show that the best model used all classes of extracted features (MFCCs, chroma, speech pauses, spectral centroid, roll-off, and fundamental frequency) but not all features (e.g., the mean of MFCC 9). According to the importances, the highest ranks have the phonetic features (spectral roll-off, seventh and eighth MFCC, and the second chromagram), followed by prosodic features (speech pauses, fundamental frequency). The feature importances of the second-best model can be found in the Appendix.

Summarized, the regression model proves beneficial for detecting unexpected shifts in subjective well-being, yielding precise numerical estimates of QOL scores, and excelling in quantitative evaluations. Given the regression's predictive power, moderate to small shifts in the subjective well-being scores can be detected, allowing the initiation of necessary interventions more quickly in case

Table 2 Regression model results

Model	Ø MAE	Ø RMSE
Random (educated) guesses	20.30 ± 1.60	24.80 ± 1.90
Support vector regression	10.90 ± 2.17	14.29 ± 2.27
Random forest regression	13.33 ± 2.84	16.64 ± 3.53
XGBoost regression	13.70 ± 2.55	17.02 ± 3.14
Lasso regression	17.44 ± 2.59	21.50 ± 3.00

Bold values indicate the best regression model results

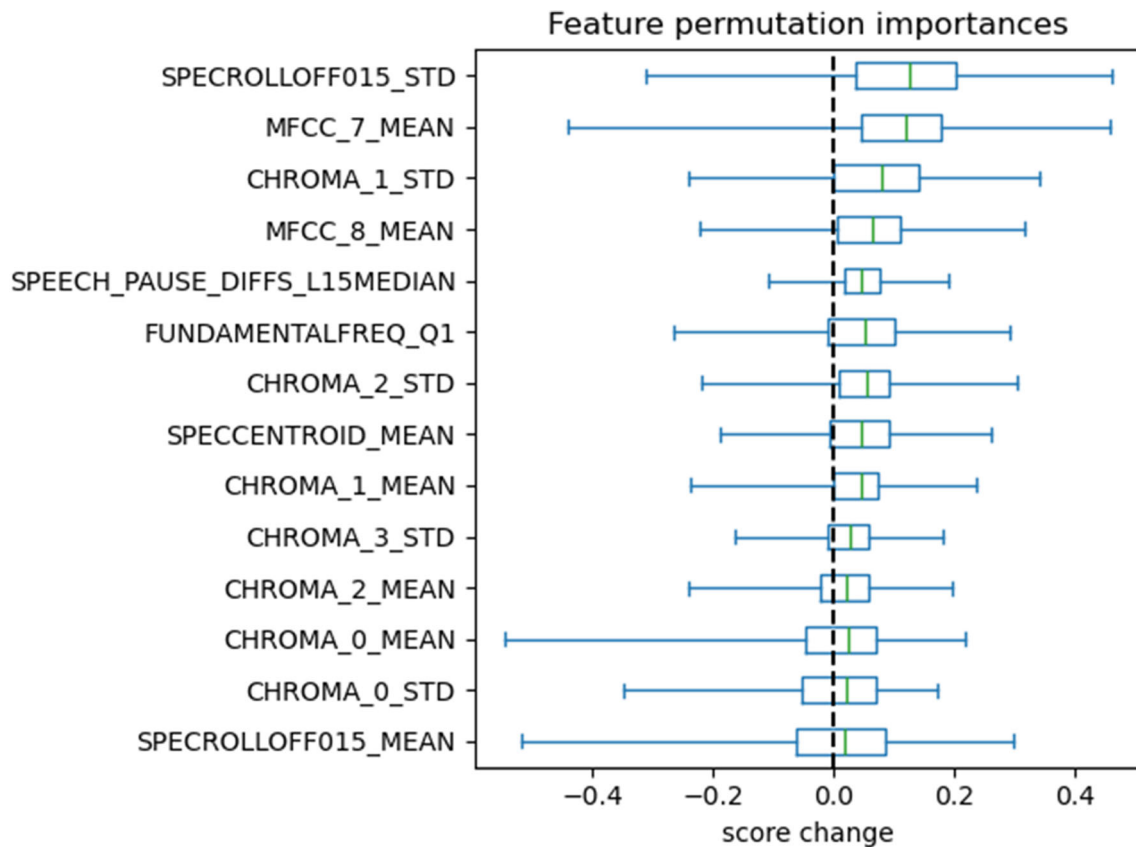


Fig. 2 Feature permutation importances of the support vector regression model

of a sudden decline in the individual's subjective well-being.

7 Discussion and Contribution

Answering the research question: the proposed ML model was found to be suitable for accurately assessing the subjective well-being of older adults from natural speech. The results offer a variety of contributions to theory and practice.

First, the model shows the capacity to predict an individual's subjective well-being – as calculated by the QOLs – yielding precise numerical estimates. This supports the proposition that the human voice (subconsciously) carries information about how the speaker feels. Acoustic phonetic and prosodic theories provided theoretical arguments for the initial proposition, and the model presented confirms it. Although previous research provides insights into the automated assessment of various human feelings and emotions via voice analysis (Ververidis and Kotropoulos 2006), the accurate measurement of subjective well-being, a complex psychological construct (Diener et al. 2009), extends our current understanding. Therefore, in a theory-testing capacity (Gregor and Klein 2014), this research

confirms the subconscious transmission of an individual's subjective well-being via natural language.

In a broader theoretical contribution, the findings lay the ground for an extended handling of the human well-being construct. As Kjell et al. (2022) discussed, human well-being is too complex to be measured accurately by closed questions in a traditional questionnaire. They argue that open questions are much better suited, although the statistical calculations become more complex. Given the advances in ML/NLP, these obstacles can be overcome. The ML model presented shows a way to assess a speaker's subjective well-being even without asking questions, purely from a person speaking, a novel step toward reimagining digital health-based tools. This reduces negative impacts on the participant (from answering the same questions repeatedly) and enables continuous monitoring as long as the participant speaks, and the ML model can capture the voice. These new possibilities open a wide variety of applications in psychological assessments, including longitudinal studies with high-frequency data collection. Due to the larger data sets per individual, longitudinal data collection will enable much finer-grained assessments of subjective well-being and the deviations towards good and bad feelings. This would result in more accurate information than those we derive from traditional

questionnaires, including the QOLs. Another benefit is the possibility of calibrating the instrument towards specific user groups (e.g., seniors living an active social life vs. seniors living in isolation).

If deployed to a larger number of participants and with a longitudinal study layout, the ML model could help research overcome the traditional obstacle of reliability of self-reported data. All instruments used today rely on the data a researcher asks the participants or the participants provide themselves. The limitations of self-reported data are well known and include social desirability response bias, i.e., the desire to respond in a way that is perceived to be beneficial to the researcher rather than authentically, and the Hawthorne effect, i.e., a potential behavior change when aware of being observed (Barata et al. 2022; Ross and Bibler Zaidi 2019). The presented ML model offers ways to overcome these limitations, as data collection can be done continuously over long periods. Over time, participants tend to forget about the algorithm, as they do not interact directly with it. The model is just a silent listener in the background, and -in the longer run- the participants will not behave differently from their everyday routines (the positive side of the Hawthorne effect).

In terms of advancements in digital health, the ML model provides numerous applications. For instance, the treatment of patients with chronic diseases if their medication needs to be adjusted. Contrary to today's typical process where the patient visits the physician, for example, weekly to report on the effect of the new medicine, the model could do this measurement more frequently, providing a time series of the patient's subjective well-being during this sensitive period until the new medication is fully adjusted. The healing process of patients outside the hospital could be monitored more closely, the physical and mental state of long-term patients (those patients who live with their illness for years) could be monitored more effectively and accurately, and new ways to assess the effectiveness of new medications (evidence-based medicine) are possible. The ML model offers more effective ways of measurement and more accurate data compared to the methods used today. Given the advances in data analytics, the numerous data points could automatically be captured, aggregated, and visualized in conjunction with peer groups of patients to benefit treating physicians and caregivers. It would enable these groups to identify increases and declines in the patient's subjective well-being and unfamiliar patterns and mood swings that may give reason to increase coverage.

From a commercial point of view, practice can apply the findings to develop innovative products and services to benefit the aging generation. It is of critical importance that societies around the world find solutions to the challenges of demographic change, leading to more mature adults and

fewer potential caregivers. The aging society implies that more and more seniors strive to live independently in their accommodations. This eases the pressure on senior homes and facilities while increasing senior citizens' general subjective well-being. However, with age comes motoric, cognitive decline, and chronic diseases. By harnessing the capabilities of the ML model in creating market-ready solutions based on the presented model, it is possible to enable more seniors to maintain their independence longer.

If the ML model is included in a solution that forms part of seniors' daily routines, it would be possible to monitor their subjective well-being continuously. This would give relatives and friends a better sense of the senior's subjective well-being, much better than the standard 'everything is good' one hears on the phone. In case of sudden drops and significant adverse developments, caregivers and physicians can take action, and relatives and friends are encouraged to intensify contact.

8 Limitations, Future Work, and Avenues for Further Research

The following limitations must be considered when assessing the ML model's capabilities: (1) The dataset used to train the algorithm comprises a comparatively small number of recordings. Although previous research demonstrates that small datasets can produce solid results (Stasak et al. 2021), a larger dataset could increase the robustness of the findings. (2) Well-being is perceived subjectively by the individual. Therefore, the limitations of self-reported data apply. Although the interviews were conducted carefully to ensure the highest level of scientific professionalism, it cannot be ruled out that the data is influenced by social desirability bias or a Hawthorne effect.

8.1 Future Work

Of the research project will focus on strengthening the algorithm and developing a standalone product for daily use: (1) Calibrate the algorithm with a larger set of training data, especially with participants speaking dialect and/or informants from other geographic regions. (2) Enhance the technical capabilities of the ML model to work 'automatically and online' (i.e., while the person is speaking) without the need to record the conversation and manually run it through the analysis cycle. (3) Develop a technology adoption approach to include the ML model into an older person's daily routines so that the voice analysis can be conducted in the most natural way of conversation. One could envision a voice assistant or an app frequently checking an individual's subjective well-being. Given the substantial increase in voice interactions on the internet

combined with the declining motoric capabilities of older adults, it seems realistic to assume that voice commands will replace the keyboard as the dominant mechanism for using devices and apps over the coming years. (4) Develop the integration of trusted caregivers (relatives, friends, physicians, professional caregivers, etc.) so that helpful interventions can be provided when necessary.

8.2 Further Research

Should focus on (1) enhancing the granularity of the predictive capabilities of the algorithm. Although the QOLs are reliable instruments, their assessment of the users' subjective well-being is comparatively coarse. With the switch from the infrequent physical questionnaire to an automated, widespread voice-based analysis comes the possibility of making more observations of the individual, which could and should result in a much finer assessment of subjective well-being than possible today. (2) It would also be promising to conduct longitudinal studies to see how subjective well-being changes over time and study those cycles in terms of duration, dynamics of change, and magnitude for continuous slow declines and sudden drops in subjective well-being. (3) The use of the WHOQOLs, adapted to numerous languages and cultural backgrounds, allows for replicating the study in other countries.

Comparing those findings to this dataset could provide valuable insights into the influence of culture, habits, etc., on the perceptions of subjective well-being of older adults in a multicultural context.

9 Conclusion

This paper presents a successfully tested ML model for assessing the subjective well-being of older adults, measured against the WHOQOL-BREF and WHOQOL-OLD questionnaires, by analyzing the human voice. The proposed model offers new insights for health monitoring and assessment by analyzing vocal attributes through NLP methods. By combining subjective well-being theories, acoustic phonetics, and prosodic theories with NLP techniques, including feature extraction from speech and regression analysis, the ML model enriches our understanding of the relationship between the human voice and the subjective well-being of the speaker.

Appendix

See (Table 3, Fig. 3).

Table 3 Feature descriptions

Feature name	Description	Category
MFCC_6	Sixth mel-frequency cepstrum coefficient which corresponds to the amplitudes at about 1147 Hz	Phonetic
MFCC_7	Seventh mel-frequency cepstrum coefficient which corresponds to the amplitudes at about 1280 Hz	Phonetic
MFCC_8	Eight mel-frequency cepstrum coefficient which corresponds to the amplitudes at about 1417 Hz	Phonetic
MFCC_9	Ninth mel-frequency cepstrum coefficient which corresponds to the amplitudes at about 1559 Hz	Phonetic
SPECROLLOFF015	Spectral roll-off for estimating the minimum spectral mass (15% roll-off percentage)	Phonetic
SPECCENTROID	Spectral centroid for estimating the center of the spectral mass	Phonetic
CHROMA_0	First chromagram out of four, corresponding to pitch class C	Acoustic
CHROMA_1	Second chromagram out of four, corresponding to pitch class D#	Acoustic
CHROMA_2	Third chromagram out of four, corresponding to pitch class G	Acoustic
CHROMA_3	Fourth chromagram out of four, corresponding to pitch class B	Acoustic
SPEECH_PAUSE_DIFFS_L15MEDIAN	Number of speech pauses divided by the length of the audio file. Speech pauses were defined as pauses larger than 1.5 times the median of the diffs between two consecutive onsets	Prosodic
FUNDAMENTALFREQ	Fundamental frequency estimation (F0) of a speaker	Prosodic

The four MFCC features were calculated from the first ten coefficients spread over a frequency range between 500 and 6000 Hz and for a sampling rate of 44.1 kHz. Various statistics were applied to the features: The mean (MFCC_6_MEAN, MFCC_7_MEAN, MFCC_8_MEAN, MFCC_9_MEAN, SPECROLLOFF015_MEAN, SPECCENTROID_MEAN, CHROMA_0_MEAN, CHROMA_1_MEAN, CHROMA_2_MEAN and CHROMA_3_MEAN), the standard deviation (SPECROLLOFF015_STD, CHROMA_0_STD, CHROMA_1_STD, CHROMA_2_STD, CHROMA_3_STD) and the first and third quantiles (FUNDAMENTALFREQ_Q1, FUNDAMENTALFREQ_Q3). Overall, with SPEECH_PAUSE_DIFFS_L15MEDIAN included, we utilized 18 features

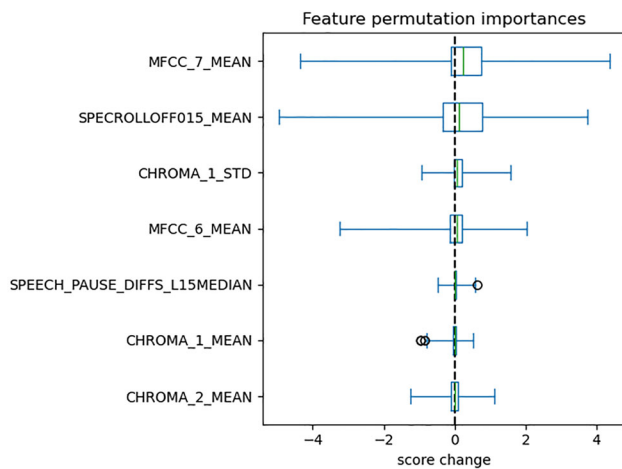


Fig. 3 Feature permutation importances of the second-best model (random forest regression). All the features that were important for the second-best model (random forest regression) coincide with the features important for the best model (support vector regression) with MFCC_6_MEAN as an exception. MFCC_7_MEAN and CHROMA_1_STD are ranked very high for both models

Acknowledgements The authors would like to thank Technology Transfer Center, Günzburg, Germany for funding the data collection for this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmad R, Siemon D, Gnewuch U, Robra-Bissantz S (2022) Designing personality-adaptive conversational agents for mental health care. *Inf Syst Front* 24(3):923–943. <https://doi.org/10.1007/s10796-022-10254-9>
- An H, Lu X, Shi D, Yuan J, Li R, Pan T (2019) Mental health detection from speech signal: a convolution neural networks approach. In: 2019 International Joint Conference on Information, Media and Engineering. IEEE, Osaka, pp 436–439. <https://doi.org/10.1109/IJCIIME49369.2019.00094>
- Artola G, Carrasco E, Rebesch KM, Larburu N, Berges I (2021) Behavioral anomaly detection system for the wellbeing assessment and lifestyle support of older people at home. *Proc Comput Sci* 192:2047–2057. <https://doi.org/10.1016/j.procs.2021.08.211>
- Barata J, da Cunha PR, de Figueiredo AD (2022) Self-reporting limitations in information systems design science research. *Bus Inf Syst Eng* 65(2):143–160. <https://doi.org/10.1007/s12599-022-00782-8>
- Barnes J, Shattuck-Hufnagel S (2022) Prosodic theory and practice. MIT Press. <https://doi.org/10.7551/mitpress/10413.001.0001>
- Beltrami D, Gagliardi G, Rossini Favretti R, Ghidoni E, Tamburini F, Calzà L (2018) Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline? *Front Aging Neurosci.* <https://doi.org/10.3389/fnagi.2018.00369>
- Bhavan A, Chauhan P, Shah RR (2019) Bagged support vector machines for emotion recognition from speech. *Knowl Based Syst.* <https://doi.org/10.1016/j.knosys.2019.104886>
- Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: Proceedings of the 5th annual workshop on Computational learning theory, Pittsburgh
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
- Statistisches Bundesamt (2023) Lebensformen älterer Menschen. Statistisches Bundesamt. <https://www.destatis.de/DE/Themen/Querschnitt/Demografischer-Wandel/Aeltere-Menschen/lebensformen.html>. Accessed 18 Dec 2023
- Byun S-W, Kim J-H, Lee S-P (2021) Multi-modal emotion recognition using speech features and text-embedding. *Appl Sci.* <https://doi.org/10.3390/app11177967>
- Centers for Disease Control Prevention (2012) Identifying vulnerable older adults and legal options for increasing their protection during all-hazards emergencies: a cross-sector guide for states and communities. U.S. Department of Health and Human Services, Atlanta
- Chachamovich E, Fleck MP, Trentini C, Power M (2008) Brazilian WHOQOL-OLD Module version: a Rasch analysis of a new instrument. *Rev Saude Publica* 42(2):308–316. <https://doi.org/10.1590/s0034-89102008000200017>
- Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. <https://doi.org/10.1145/2939672.2939785>
- Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comp Sci* 7:e623
- Conrad I, Matschinger H, Riedel-Heller S, von Gottberg C, Kilian R (2014) The psychometric properties of the German version of the WHOQOL-OLD in the German population aged 60 and older. *Health Qual Life Outcomes.* <https://doi.org/10.1186/s12955-014-0105-4>
- Conrad I, Matschinger H, Kilian R, Riedel-Heller SG (2016) WHOQOL-OLD und WHOQOL-BREF: Handbuch für die deutschsprachigen Versionen der WHO-Instrumente zur Erfassung der Lebensqualität im Alter. Hogrefe, Göttingen
- Cooke PJ, Melchert TP, Connor K (2016) Measuring well-being: a review of instruments. *Couns Psychol* 44(5):730–757. <https://doi.org/10.1177/0011000016633507>
- Corrales-Astorgano M, Martínez-Castilla P, Escudero-Mancebo D, Aguilar L, González-Ferreras C, Cardeñoso-Payo V (2019) Automatic assessment of prosodic quality in down syndrome: analysis of the impact of speaker heterogeneity. *Appl Sci.* <https://doi.org/10.3390/app9071440>
- Czaja SJ, Ceruso M (2022) The promise of artificial intelligence in supporting an aging population. *J Cogn Eng Decis Mak* 16(4):182–193. <https://doi.org/10.1177/15553434221129914>
- de Cheveigne A, Kawahara H (2002) YIN, a fundamental frequency estimator for speech and music. *J Acoust Soc Am* 111(4):1917–1930. <https://doi.org/10.1121/1.1458024>
- DeSouza DD, Robin J, Gumus M, Yeung A (2021) Natural language processing as an emerging tool to detect late-life depression. *Front Psychiatry.* <https://doi.org/10.3389/fpsy.2021.719125>

- Diener E (1984) Subjective well-being. *Psychol Bull* 95(3):542–575. <https://doi.org/10.1037/0033-2909.95.3.542>
- Diener E, Lucas RE, Oishi S (2009) Subjective well-being: the science of happiness and life satisfaction. In: *The oxford handbook of positive psychology*, 2nd edn, Oxford University Press, New York, pp 187–194
- Diex J (2019) Perceived needs of elderly for living a self-reliant life: implications for municipal health policy. *Eur J Publ Health*. <https://doi.org/10.1093/eurpub/ckz186.481>
- Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V (1996) Support vector regression machines. In: *Proceedings of the 9th International Conference on Neural Information Processing Systems*, Denver
- Flake R, Kochskämper S, Risius P, Seyda S (2018) Fachkräfteengpass in der Altenpflege. *IW-Trends*, vol 3. <https://hdl.handle.net/10419/194600>
- Gaertner B, Scheidt-Nave C, Koschollek C, Fuchs J (2023) Gesundheitliche Lage älterer und hochaltriger Menschen in Deutschland: Ergebnisse der Studie Gesundheit 65+. *J Health Monit* 8(3):7–31. <https://doi.org/10.25646/11564>
- Godino-Llorente JI, Gomez-Vilda P (2004) Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Trans Biomed Eng* 51(2):380–384. <https://doi.org/10.1109/TBME.2003.820386>
- Gregor S, Klein G (2014) Eight obstacles to overcome in the theory testing genre. *J Assoc Inf Syst* 15(11):I–XIX. <https://doi.org/10.17705/1jais.00382>
- Gupta MV, Vaikole S, Oza AD, Patel A, Burduhos-Nergis DP, Burduhos-Nergis DD (2022) Audio-visual stress classification using cascaded RNN-LSTM networks. *Bioengineering* 9(10):510. <https://doi.org/10.3390/bioengineering9100510>
- Hubbard DJ, Faso DJ, Assmann PF, Sasson NJ (2017) Production and perception of emotional prosody by adults with autism spectrum disorder. *Autism Res* 10(12):1991–2001. <https://doi.org/10.1002/aur.1847>
- Huppert FA, So TT (2013) Flourishing across Europe: application of a new conceptual framework for defining well-being. *Soc Indic Res* 110(3):837–861. <https://doi.org/10.1007/s11205-011-9966-7>
- Keyes CL (2005) Mental illness and/or mental health? Investigating axioms of the complete state model of health. *J Consult Clin Psychol* 73(3):539–548. <https://doi.org/10.1037/0022-006X.73.3.539>
- Khanbhai M, Anyadi P, Symons J, Flott K, Darzi A, Mayer E (2021) Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review. *BMJ Health Care Inform*. <https://doi.org/10.1136/bmjhci-2020-100262>
- Khodabakhsh A, Yesil F, Guner E, Demiroglu C (2015) Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech. *EURASIP J Audio Speech Music Proc* 2015:1–15. <https://doi.org/10.1186/s13636-015-0052-y>
- Kim S, Kwon N, O'Connell H (2019) Toward estimating personal well-being using voice. *arXiv:1910.10082*
- Kjell ONE, Sikström S, Katarina Kjell H, Schwartz A (2022) Natural language analyzed with AI-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy. *Sci Rep*. <https://doi.org/10.1038/s41598-022-07520-w>
- Klapuri A, Davy M (2006) *Signal processing methods for music transcription*. Springer
- Ladefoged P, Johnson K (2014) *Articulation and acoustics*. In: *A course in phonetics*, 7th edn. Cengage Learning, pp. 2–32
- Lent RW (2004) Toward a unifying theoretical and practical perspective on well-being and psychosocial adjustment. *J Counsel Psychol* 51(4):482–509. <https://doi.org/10.1037/0022-0167.51.4.482>
- Li Y, Jiang Y, Tian D, Hu L, Lu H, Yuan Z (2019) AI-enabled emotion communication. *IEEE Netw* 33(6):15–21. <https://doi.org/10.1109/MNET.001.1900070>
- Lin L, Chen X, Shen Y, Zhang L (2020) Towards automatic depression detection: a BiLSTM/1D CNN-Based Model. *Appl Sci* 10(23):1–20. <https://doi.org/10.3390/app10238701>
- Logan B (2000) Mel frequency cepstral coefficients for music modeling. In: *International society for music information retrieval conference*. Plymouth. https://ismir2000.ismir.net/papers/logan_abs.pdf
- Lucas-Carrasco R (2012) The WHO quality of life (WHOQOL) questionnaire: Spanish development and validation studies. *Qual Life Int J Qual Life Asp Treatm Care Rehab* 21(1):161–165. <https://doi.org/10.1007/s11136-011-9926-3>
- Martinho D, Carneiro J, Novais P, Neves J, Corchado J, Marreiros G (2019) A conceptual approach to enhance the well-being of elderly people. In: *Oliveira PM, Novais P, Reis LP (eds) Progress in Artificial Intelligence: 19th EPIA Conference on Artificial Intelligence*, Vila Real. Springer, Cham, pp 50–61. https://doi.org/10.1007/978-3-030-30244-3_5
- McFee B, Raffel C, Liang D, Ellis DPW, McVicar M, Battenberg E, Nieto O (2015) librosa: Audio and music signal analysis in Python. In: *Proceedings of the 14th Python in Science Conference*
- Nakagawa S, Enomoto D, Yonekura S, Kanazawa H, Kuniyoshi Y (2020) New telecare approach based on 3D convolutional neural network for estimating quality of life. *Neurocomput* 397:464–476. <https://doi.org/10.1016/j.neucom.2019.09.112>
- National Institute on Aging (2021) Aging and health: trends and statistics. <https://www.nia.nih.gov/research/dbsr/data-resources-behavioral-and-social-research-aging>. Accessed 26 May 2023
- Pentari A, Kafentzis G, Tsiknakis M (2024) Speech emotion recognition via graph-based representations. *Sci Rep*. <https://doi.org/10.1038/s41598-024-52989-2>
- Perez M, Jin W, Le D, Carlozzi N, Dayalu P, Roberts A, Provost EM (2018) Classification of Huntington Disease using acoustic and lexical features. In: *Interspeech, Hyderabad*, pp 1898–1902. <https://doi.org/10.21437/interspeech.2018-2029>
- Power M, Quinn K, Schmidt S, Whoqol-Old Group (2005) Development of the WHOQOL-old module. *Qual Life Res* 14:2197–2214
- Rathina XA, Mehata K, Ponnavaikko M (2012) Basic analysis on prosodic features in emotional speech. *Int J Compu Sci Eng Appl* 2(4):99–107. <https://doi.org/10.5121/ijcsea.2012.2410>
- Reiner BI (2013) Expanding the functionality of speech recognition in radiology: creating a real-time methodology for measurement and analysis of occupational stress and fatigue. *J Digit Imaging* 26(1):5–9. <https://doi.org/10.1007/s10278-012-9540-0>
- Rejaibi E, Komaty A, Meriaudeau F, Agrebi S, Othmani A (2022) MFCC-based recurrent neural network for automatic clinical depression recognition and assessment from speech. *Biomed Signal Process Control* 71:103107. <https://doi.org/10.1016/j.bspc.2021.103107>
- Ribeiro O, Araújo L, Figueiredo D, Paúl C, Teixeira L (2021) The caregiver support ratio in Europe: estimating the future of potentially (un)available caregivers. *Healthcare* 10(1):11. <https://doi.org/10.3390/healthcare10010011>
- Robert Ladd D (2008) *Intonational phonology*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511808814>
- Ross PT, Bibler Zaidi NL (2019) Limited by our limitations. *Perspect Med Educ* 8:261–264. <https://doi.org/10.1007/s40037-019-00530-x>
- Rusz J, Cmejla R, Ruzickova H, Ruzicka E (2011) Quantitative acoustic measurements for characterization of speech and voice

- disorders in early untreated Parkinson's disease. *J Acoust Soc Am* 129(1):350–367. <https://doi.org/10.1121/1.3514381>
- Ryff CD (2014) Psychological well-being revisited: advances in the science and practice of eudaimonia. *Psychother Psychosom* 83(1):10–28. <https://doi.org/10.1159/000353263>
- Sanchez MH, Vergyri D, Ferrer L, Richey C, Garcia P, Knoth B, Jarrold W (2011) Using prosodic and spectral features in detecting depression in elderly males. In: Annual Conference of the International Speech Communication Association, Florence, pp 3001–3004. <https://doi.org/10.21437/Interspeech.2011-751>
- Schuller BW, Batliner A, Bergler C, Messner E-M, Hamilton A, Amiriparian S, Baird A, Rizo G, Schmitt M, Stappen L (2020) The INTERSPEECH 2020 Computational Paralinguistics Challenge: elderly emotion, breathing & masks. In: Proceedings of the Interspeech 2020, Shanghai
- Skevington SM, McCrate FM (2012) Expecting a good quality of life in health: assessing people with diverse diseases and conditions using the WHOQOL-BREF. *Health Expect* 15(1):49–62. <https://doi.org/10.1111/j.1369-7625.2010.00650.x>
- Stasak B, Huang Z, Razavi S, Joachim D, Epps J (2021) Automatic detection of COVID-19 based on short-duration acoustic smartphone speech analysis. *J Healthcare Inform Res* 5:201–217. <https://doi.org/10.1007/s41666-020-00090-4>
- Suresh C, Sathvik MC, Deepthi N, Purnima KMS, Chouhan KPS (2023) A study on cross-lingual speech emotion analysis using natural language processing. In: International Conference on Sustainable Computing and Data Communication Systems, Erode, IEEE, pp 808–815
- Tariq Z, Shah SK, Lee Y (2019) Speech emotion detection using IoT based deep learning for health care. In: IEEE International Conference on Big Data, Los Angeles, pp 4191–4196. <https://doi.org/10.1109/BigData47090.2019.9005638>
- The WHOQOL Group (1998a) Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med* 28(3):551–558. <https://doi.org/10.1017/S0033291798006667>
- The WHOQOL Group (1998b) The World Health Organization quality of life assessment (WHOQOL): development and general psychometric properties. *Soc Sci Med* 46(12):1569–1585. [https://doi.org/10.1016/s0277-9536\(98\)00009-4](https://doi.org/10.1016/s0277-9536(98)00009-4)
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc B (methodological)* 58(1):267–288
- Tibshirani R (2011) Regression shrinkage and selection via the lasso: a retrospective. *J R Stat Soc B* 73:273–282. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>
- United Nations (2019) World population ageing 2019: highlights. United Nations, Department of Economic and Social Affairs, New York
- Ververidis D, Kotropoulos C (2006) Emotional speech recognition: resources, features, and methods. *Speech Commun* 48(9):1162–1181. <https://doi.org/10.1016/j.specom.2006.04.003>
- Vik MH, Carlquist E (2018) Measuring subjective well-being for policy purposes: the example of well-being indicators in the WHO “Health 2020” framework. *Scand J Public Health* 46(2):279–286. <https://doi.org/10.1177/1403494817724952>
- Wang K, An N, Li BN, Zhang Y, Li L (2015) Speech emotion recognition using Fourier parameters. *IEEE Trans Affect Comput* 6(1):69–75. <https://doi.org/10.1109/TAFFC.2015.2392101>
- Weed E, Fusaroli R (2020) Acoustic measures of prosody in right-hemisphere damage: a systematic review and meta-analysis. *J Speech Lang Hear Res* 63(6):1762–1775. https://doi.org/10.1044/2020_JSLHR-19-00241
- World Health Organization (2012) The World Health Organization Quality of Life (WHOQOL). <https://www.who.int/publications/i/item/WHO-HIS-HSI-Rev.2012.03>
- World Health Organization (2020) Constitution of the world health organization. World Health Organization. <https://www.who.int/about/accountability/governance/constitution>. Accessed 9 Nov 2023
- World Health Organization (2023) Ageing and health. World Health Organization,. <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>. Accessed 10 Nov 2023
- Wu P, Wang R, Lin H, Zhang F, Tu J, Sun M (2023) Automatic depression recognition by intelligent speech signal processing: a systematic survey. *CAAI Trans Intell Technol* 8(3):701–711. <https://doi.org/10.1049/cit2.12113>
- Yalamanchili B, Kota NS, Abbaraju MS, Nadella VSS, Alluri SV (2020) Real-time acoustic based depression detection using machine learning techniques. In: International Conference on Emerging Trends in Information Technology and Engineering, Vellore, IEEE. <https://doi.org/10.1109/ic-ETITE47903.2020.394>
- Yıldırım M, Çelik Tanrıverdi F (2020) Social support, resilience and subjective well-being in college students. *J Posit School Psychol* 5(2):127–135. <https://doi.org/10.47602/jpsp.v5i2.229>
- Zunic A, Corcoran P, Spasic I (2020) Sentiment analysis in health and well-being: systematic review. *JMIR Med Inf* 8(1):e16023. <https://doi.org/10.2196/16023>