Wearable Technology and Deep Learning for Burnout Prediction

Samady Ahmed, Chibani Fahd,
IT Department, FSTT-UAE,

***Abstract*—Burnout is a critical issue affecting shift workers, who are particularly vulnerable due to the high demands and irregular schedules inherent in their professions. It is characterized by emotional exhaustion, depersonalization, and reduced personal accomplishment, with profound implications for mental health and workplace productivity. Early detection of burnout is vital to prevent its progression and mitigate its impact. This study aims to harness wearable technology, such as the Empatica E4 device, and machine learning to predict burnout risk by analyzing physiological data alongside synthetic metrics. Data collection involved continuous monitoring of electrodermal activity, heart rate, and skin temperature, complemented by synthetic features modeling factors like mood, work stress, quality of life, and depressive symptoms. A convolutional neural network (CNN) was developed to analyze these inputs, achieving moderate performance metrics, including a root mean square error (RMSE) of 0.065 and an R^2 score of 0.23. Despite limitations in the dataset, which included a reliance on synthetic data and a small sample size, the study demonstrates the potential of integrating wearable sensor data and machine learning for burnout detection. These findings emphasize the need for richer datasets and real-world validations but highlight the promise of such technologies in enabling early interventions. By improving burnout risk assessment, this research offers a pathway to personalized stress management strategies and workplace policies that promote mental health, reduce turnover, and enhance productivity, particularly for shift workers in high-stress environments.**

***Index Terms*—Deep Learning, Wearable Technology, Burnout, CNNs.**

I. Introduction

BURNOUT is a pervasive occupational phenomenon, particularly among shift workers, who are disproportionately affected due to the irregularity of their work schedules, high job demands, and limited recovery periods. This condition, characterized by emotional exhaustion, depersonalization, and diminished personal accomplishment, has far-reaching consequences, not only for individual mental and physical health but also for organizational performance and societal well-being. Despite its recognition as a workplace syndrome by the World Health Organization (WHO), the accurate prediction and timely detection of burnout remain significant challenges, especially in populations with variable work patterns like shift workers.

Traditional methods for assessing burnout, such as self-reported surveys and observational studies, are limited by their subjective nature, susceptibility to recall bias, and inability to provide continuous, real-time insights. Furthermore, the complex interplay of physiological, psychological, and environmental factors contributing to burnout makes it challenging to develop comprehensive models for risk prediction. For shift workers, additional complications arise from circadian rhythm disruptions, irregular physical activity levels, and varying stress triggers across different occupational contexts. Addressing these issues requires innovative solutions that can capture the multidimensional nature of burnout while accommodating the dynamic and often unpredictable routines of shift workers.

Wearable technology and deep learning offer a transformative approach to overcoming these challenges. Wearable devices, such as the Empatica E4, provide an unprecedented capability to collect continuous, objective physiological data, including electrodermal activity (EDA), heart rate (HR), and skin temperature (ST). These metrics can serve as proxies for stress levels and recovery states, enabling real-time monitoring of physiological responses to occupational stressors. When combined with synthetic features that model psychological and work-related dimensions, such as mood, workload, and quality of life, wearable technology can provide a holistic view of burnout risk factors.

The integration of deep learning models further enhances this approach by leveraging their ability to identify intricate patterns and relationships within high-dimensional data. Unlike traditional statistical methods, deep learning techniques such as convolutional neural networks (CNNs) excel at capturing temporal and spatial dependencies in time-series data, making them particularly suitable for analyzing physiological signals. Additionally, counterfactual explanations and other interpretability techniques can bridge the gap between model accuracy and practical applicability, ensuring that insights derived from these models are actionable and aligned with real-world burnout management strategies.

This paper aims to address the challenges of burnout prediction in shift workers by exploring the combined potential of wearable technology and deep learning. By leveraging physiological data, synthetic metrics, and advanced machine learning techniques, this study seeks to develop a predictive framework that not only identifies individuals at risk of burnout but also provides actionable insights for targeted interventions. In doing so, it contributes to the broader goal of improving mental health outcomes and fostering sustainable workplace practices in high-stress occupational settings.

II. Literature Review

Burnout, recognized by the WHO as a workplace stress syndrome, is a widespread challenge in healthcare, particularly

MST-IASD, November the 28th, 2024

among nurses. This issue is exacerbated during high-demand periods, such as the COVID-19 pandemic, where nurses face heavy workloads, emotional strain, and heightened exposure to health risks. Prolonged stress can lead to significant health consequences, including hypertension, obesity, and mental health disorders [1], highlighting the importance of continuous monitoring and interventions to enhance workplace well-being.

The integration of machine learning (ML) with wearable devices has revolutionized stress and burnout detection. Devices such as Empatica E4 wristbands and Garmin smartwatches capture continuous physiological data [1], including electrodermal activity (EDA), heart rate (HR), skin temperature (ST), and accelerometer readings, enabling real-time assessment of stress in the workplace. ML algorithms like random forests and support vector machines analyze these data streams, focusing on features such as variability, mean, and maximum values, to detect stress with high accuracy [2].

Recent studies have demonstrated the utility of these technologies. For instance, a dataset collected from 15 female emergency room nurses during the COVID-19 pandemic recorded over 1,250 hours of physiological data, with 171 hours labeled as stressful using self-reported surveys and physiological indicators. Stress contributors, including workload and patient interactions, were identified through post-shift surveys. Among the physiological signals, EDA showed a strong correlation with stress, while signals like HR were more influenced by physical activity, reducing their reliability as stress indicators.

The BROWNIE study [3] further validated the potential of combining wearable technology with workplace and psychological metrics for burnout prediction. Using a decentralized approach to minimize participant burden, the study integrated physiological data from smartwatches with psychological and workplace surveys. This multi-modal data approach has been instrumental in enhancing the accuracy of burnout prediction models.

Machine learning has played a pivotal role in advancing burnout detection. Features such as EDA variability and rhythmic patterns in physiological data have proven critical in distinguishing stress levels across different occupational contexts. Studies like the AMED dataset [4] have underscored the importance of rhythm features and multi-modal data analysis. Clustering algorithms, such as k-means, have further enabled the identification of burnout subtypes, allowing for targeted interventions tailored to specific stress profiles.

The combination of wearable technology and ML offers a promising framework for real-time burnout detection, early interventions, and workplace policy improvements. The data collected can guide tailored interventions for high-stress environments, benefiting both healthcare workers and patients. However, challenges remain, including the complexities of data labeling, the influence of confounding variables like physical activity, and recall biases in self-reported surveys, necessitating further research to optimize these systems.

III. Data Description:

A. Primary Features

The dataset utilized in this study spans a period of 3,190 days, with a daily frequency after preprocessing. Originally, the data was captured at a high frequency (every 31 milliseconds) and subsequently resampled to intervals of 2 minutes. For the variables Electrodermal Activity (EDA), Heart Rate (HR), and body Temperature (TEMP), the mean and standard deviation were calculated within these intervals, yielding the following primary features:

- **EDA_mean** and **EDA_std**: The mean and standard deviation of Electrodermal Activity.
- **HR_mean**: The mean of Heart Rate.
- **TEMP_mean**: The mean of body Temperature.
- **prev_day_HR_mean** and **prev_day_EDA_mean**: Capturing the lagged influence of HR and EDA from the previous day, as these factors are hypothesized to impact subsequent daily patterns.

B. Synthetic Features

Given the scarcity of open datasets that comprehensively capture both physiological and psychological features, synthetic data was generated to simulate realistic behavioral and psychological patterns. This approach enabled the creation of a dataset reflecting the complex interplay between physiological responses and psychological states, crucial for burnout prediction. However, the reliance on synthetic data presents challenges, as it may not perfectly mirror real-world scenarios, potentially introducing biases. Efforts were made to enhance the realism of the synthetic data by employing validated psychological models and expert knowledge, aiming to approximate plausible relationships between variables. Despite these measures, the inability to validate the synthetic data against real-world data underscores the need for future research to prioritize the collection and sharing of comprehensive datasets. These features are described below:

1) Mood and Anxiety:

$$\begin{aligned} mood \;&=\; base\_mood - \frac{HR_{mean}}{100} - \frac{EDA_{mean}}{10} \\ &\quad - 0.2 \cdot workload + \mathcal{N}(0, 0.5) \end{aligned}$$

Lower HR and EDA values are assumed to correlate with better mood and lower anxiety scores [5].

2) Work and Stress Metrics:

a) Work Hours Relation:

$$work\_hours \;=\; \begin{cases} 8, & \text{in workday} \\ 0, & \text{in week-ends and holidays} \end{cases}$$

b) Workload Relation:

$$\begin{aligned} workload = \;& base\_workload + 0.1 \cdot day\_index \\ & + \mathcal{N}(0, 1) \end{aligned}$$

Workload incorporates a daily trend representing increasing demand, and is higher on workdays compared to holidays [6].

3) Quality of Life and Stress Factors:

$$quality\_of\_life = 0.15 \cdot mood\_rating \\ - 0.1 \cdot anxiety\_score + \mathcal{N}(0, 0.5)$$

$$\begin{aligned} family\_stress &\sim \mathcal{N}(5, 2) \\ relationship\_satisfaction &\sim \mathcal{N}(7, 1.5) \\ financial\_stress &\sim \mathcal{N}(4, 1.5) \\ social\_support &\sim \mathcal{N}(6, 1.5) \\ recent\_life\_changes &\sim Bernoulli(0.2) \end{aligned}$$

These metrics quantify various dimensions of life quality and stressors, with realistic distributions applied for synthetic generation.

4) Depressive Symptoms:

$$\begin{aligned} &depressive\_symptoms = \; -0.1 \cdot mood\_rating \\ &+ \; 0.2 \cdot anxiety\_score \; + \; 0.15 \cdot family\_stress \\ &- \; 0.1 \cdot relationship\_satisf \; + \; 0.1 \cdot financial\_stress \\ &- \; 0.15 \cdot social\_supportquad \; + 0.3 \cdot recent\_life\_changes \\ &+ \; \mathcal{N}(0, 0.5) \end{aligned}$$

Depressive symptoms are modeled as a weighted combination of mood, anxiety, and various stressors [7].

5) Burnout Risk:

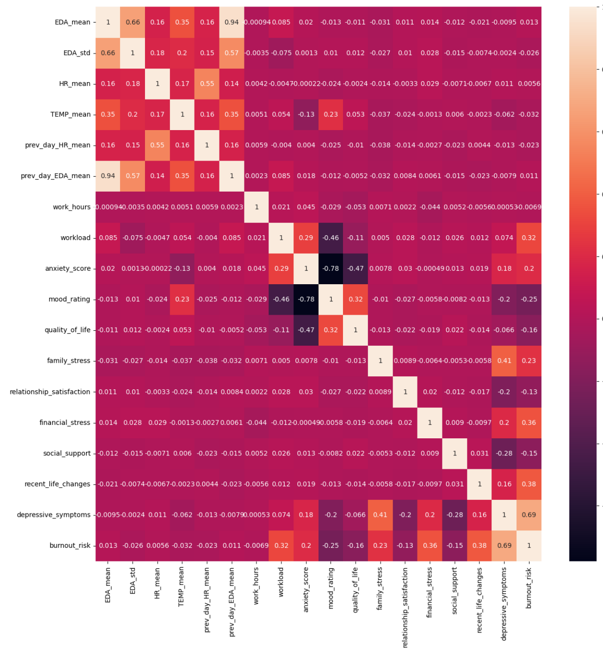
$$\begin{aligned} &z = \; -1.5 + 0.2 \cdot depressive\_symptoms\_z \\ &- 0.1 \cdot quality\_of\_life\_z + 0.1 \cdot workload\_z \\ &+ 0.1 \cdot financial\_stress\_z + 0.1 \cdot recent\_life\_changes\_z \\ &+ \mathcal{N}(0, 0.2) \end{aligned}$$

$$burnout\_risk \;\; = \;\; \frac{1}{1 + \exp(-z)}$$

Burnout risk is computed using a logistic function, incorporating scaled (z-score) variables for depressive symptoms, quality of life, workload, financial stress, and recent life changes [8]. This comprehensive feature set, derived from both empirical data and synthetic assumptions, enables an in-depth analysis of burnout risk factors, leveraging both physiological and psychological dimensions.

The Figure 1 illustrates the correlation matrix of the primary and synthetic features:

Fig. 1. Correlation Matrix of Features



IV. Methodology & Modeling

A. Data Preprocessing

The dataset was preprocessed to ensure data quality and consistency. Since most of the features are synthetic except for EDA, HR, and TEMP, the primary focus was on handling outliers, and normalizing the data. Outliers were detected using the Z-score method, with a threshold of 0.3 rolling mean and standard deviation of the previous 30 days. The data was then standardized using the z-score method to ensure that all features had a mean of 0 and a standard deviation of 1. We also had to drop a couple of features that were highly correlated with others to avoid multicollinearity issues. The dataset was split into training, validation, and test sets, with a 80-10-10 ratio, and since the data was time-series, the training set was not shuffled to preserve temporal patterns.

B. Modeling Approach

To predict burnout risk in shift workers, we employed a deep learning approach, leveraging convolutional neural networks (CNNs), the CNN-1D model architecture consisted of a single convolutional layer with 28 filters, followed by a max-pooling layer, a flatten layer, and two dense layers, with 16 units each, and a dropout rate of 0.5 and 0.2, respectively. The model was trained using the AdamW optimizer with a learning rate of 0.001, a batch size of 32, and a Mean Squared Error (MSE) loss function. The following table represents the model's general architecture:

TABLE I
CNN-1D Model Architecture

Layer	Type	Filters	Units	Activation
1	Conv1D	28	-	ReLU
2	MaxPooling1D	-	-	-
3	Flatten	-	-	-
4	Dense	-	16	Linear
5	Dropout	-	-	-
6	Dense	-	16	Linear
7	Dropout	-	-	-
8	Dense	-	1	ReLU

C. Model Evaluation

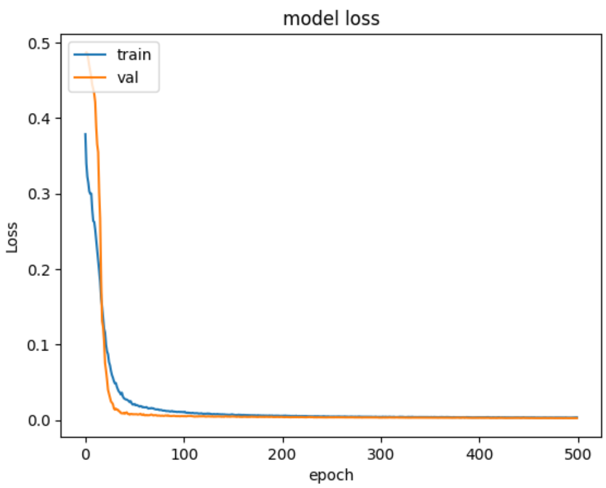
The model was evaluated using a number of metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2) score. These metrics provide insights into the model's performance, highlighting its predictive accuracy and generalization capabilities.

V. Results

After transitioning from a CNN-1D to more complex architectures like LSTM, ConvLSTM-1D, and CNN-LSTM, and after training for a 500 epochs, with early stopping based on the validation's R2 to prevent overfitting. The model initially showed promising results as evidenced by the decreasing loss curves for both training and validation sets. However, upon evaluating the model using metrics such as R^2, MSE, RMSE, and MAE, the performance was found to be suboptimal. This discrepancy suggests that while the model was effectively minimizing loss during training, it did not generalize well to unseen data, indicating potential overfitting or an inability to capture the underlying patterns in the data that translate to meaningful predictions.

The Figure 2 illustrates the model's loss during training and validation:

Fig. 2. Model Loss Plot



The model's performance in forecasting burnout risk yielded mixed results. While the RMSE of 0.065 indicates a reasonable level of accuracy given the output range of 0 to 1, the R^2 score of 0.23 reflects limited success in capturing the variance in burnout risk. This suggests that the model struggles to fully capture the complex relationships between features and burnout, likely due to the use of a limited dataset with synthetic features. These results highlight the challenges of predicting burnout and underscore the need for richer, real-world data and more advanced feature engineering to improve forecasting accuracy.

TABLE II
Model Evaluation Metrics

Metric	Value
R^2	0.23
MSE	0.0042
RMSE	0.065
MAE	0.045

VI. Discussion: Limitations and Implications

The current study aimed to predict burnout risk in shift workers using a combination of physiological data from wearable devices and synthetic features generated to simulate various aspects of mental health and work environment. However, several limitations in the dataset significantly impacted the model's ability to generalize and predict burnout risk effectively.

A. Synthetic Data Generation

A substantial portion of the dataset was comprised of synthetic data, generated based on simplified assumptions and mathematical models. While these synthetic features aimed to capture aspects such as mood, anxiety, and work stress, they were derived from equations that may not fully encapsulate the nuanced and dynamic interactions of these variables in real-world scenarios. For example, stress levels might interact with mood differently based on individual coping mechanisms or workplace support systems, complexities that simplified models often overlook.

This inherent simplification likely introduced biases and inaccuracies in the data, making it challenging for the model to learn meaningful patterns. To mitigate this issue, future research could focus on improving the realism of synthetic data by leveraging validated psychological frameworks, such as the diathesis-stress model, which accounts for individual susceptibility to stressors. Additionally, semi-supervised learning techniques could be employed to combine synthetic data with small, high-quality real-world datasets, enhancing the overall reliability of the data. Cross-validation using external datasets is also essential to ensure that the synthetic data aligns closely with real-world patterns.

B. Limited Sample Size and Homogeneity

The dataset was collected from a relatively small and homogeneous sample of 15 female nurses, which poses significant challenges for generalization. This limited demographic diversity restricts the model's applicability to broader populations of shift workers, such as those in different occupations, age groups, or geographic regions. Moreover, the seven-month data collection period may not capture long-term patterns and variations in burnout risk, such as those influenced by seasonal changes, workload fluctuations, or life events.

To address these limitations, future studies could adopt a multi-center approach to recruit a more diverse participant pool, encompassing various genders, occupations, and cultural contexts. Increasing sample size would also enable the use of advanced machine learning techniques, such as transfer learning, where models pre-trained on larger, generalized datasets (e.g., public physiological datasets) can be fine-tuned for specific populations. Expanding the data collection period to cover at least one full year could help capture the cyclical and temporal factors contributing to burnout.

C. Confounding Variables and Data Quality

Physiological data collected from wearable devices, such as heart rate (HR), was influenced by multiple confounding variables that were not fully accounted for during data collection and preprocessing. For instance, HR is affected

by factors like physical exertion, caffeine consumption, and sleep quality, all of which can introduce noise and variability into the dataset. Without proper control or adjustment for these confounders, the model may struggle to discern the true relationships between physiological signals and burnout risk.
Future research could benefit from more rigorous data preprocessing, including feature engineering to isolate stress-related HR variations from those caused by unrelated factors. Collecting additional contextual data, such as physical activity levels, dietary habits, and sleep patterns, could also help disentangle these confounding influences. Integrating advanced statistical techniques, such as causal inference models, may further aid in identifying and accounting for the true predictors of burnout.

D. Lack of Real-World Context

The dataset lacked critical real-world context, such as specific work environments, organizational policies, and social support systems, which are key determinants of burnout risk. Burnout is a multifaceted phenomenon influenced by both individual and systemic factors. For example, workplace dynamics, team relationships, and access to mental health resources significantly shape how stress and workload are perceived and managed.
To overcome this limitation, future research should consider incorporating contextual data, such as workplace culture surveys, organizational policies, and team dynamics assessments. These variables could be collected through complementary qualitative methods, such as interviews or focus groups, providing a richer understanding of the burnout phenomenon. Additionally, context-aware machine learning models, which integrate structured physiological data with unstructured contextual information, could enhance predictive accuracy and practical applicability.

VII. Future Directions

A. Environmental and Behavioral Factors

To enhance the comprehensiveness of burnout prediction models, future research should consider integrating work-specific environmental and behavioral data. Environmental factors, such as workload, job control, organizational support, work-life balance, and exposure to shift work, can provide insights into how work environment characteristics influence stress and burnout. Behavioral data, including sleep patterns, physical activity, and social interactions, can offer a deeper understanding of how lifestyle factors interact with work-related stress. Incorporating these diverse data types will likely improve the model's predictive accuracy by capturing a broader spectrum of influences on mental health.

B. Continuous Data Collection

Implementing continuous and high-frequency data collection on a daily basis would facilitate the detection of more subtle and dynamic patterns in stress and burnout development. Incorporating advanced sensor technologies and emerging more wearables could also provide richer, multi-modal datasets, such as combining physiological, behavioral, and environmental data streams.

C. Model Interpretability

While deep learning models effectively predict burnout, their lack of interpretability poses challenges, particularly in healthcare. By integrating explainable AI (XAI), we can enhance model transparency, identifying key factors like mood and work stress that influence burnout risk. This enables targeted interventions and ensures ethical considerations, such as accountability and trust, are met, fostering stakeholder adoption.

D. Personalized Interventions

Developing a feedback mechanism that connects predictions to real-world applications, such as adaptive interventions or personalized stress management plans, would make these systems more practical and impactful. Collaborating with social scientists, psychologists, and workplace managers to design and evaluate these applications will ensure that the findings translate into tangible improvements in employee well-being, productivity, and organizational health.

VIII. Conclusion

This study explored the prediction of burnout risk in shift workers by leveraging wearable sensor data and a comprehensive set of metrics derived from 15 female nurses working regular shifts in a hospital setting. The dataset, initially collected over seven months, was augmented to simulate an extended period of 8 years and 9 months, enhancing the analysis. We employed advanced deep learning techniques, particularly convolutional neural networks (CNNs), to capture temporal patterns and relationships, thereby enabling burnout prediction in shift workers.

The integration of physiological data from wearable devices, such as electrodermal activity, heart rate, and skin temperature, significantly strengthened our predictive models. Beyond these physiological measures, we incorporated a broad range of metrics, including mood, work and stress indicators, quality of life factors, depressive symptoms, and overall burnout risk, providing a multidimensional framework to explore the complex dynamics of burnout.

Despite the dataset's focus on a specific group of nurses, which limits generalizability, we addressed this limitation by generating synthetic data to simulate various aspects of mental health and work environment. These additional metrics enhanced the dataset's versatility and broadened its applicability. However, future studies should aim to validate these findings across diverse populations and settings to ensure broader relevance.

Our research underscores the potential of wearable technologies and advanced machine learning models in monitoring and mitigating burnout in high-stress occupations. By integrating physiological data with contextual metrics, this study lays a foundation for developing data-driven strategies to prevent burnout and support mental health in demanding work environments. Nonetheless, the challenge of achieving high model accuracy remains, highlighting the need for further research to refine and enhance the model's predictive capabilities.

References

[1] Seyedmajid Hosseini, Raju Gottumukkala, Satya Katragadda, Ravi Teja Bhupatiraju, Ziad Ashkar, Christoph W. Borst, and Kenneth Cochran. A multimodal sensor dataset for continuous stress detection of nurses in a hospital. volume 9, June 2022.

[2] Malgorzata Grzadzielewska. Using machine learning in burnout prediction: A survey. In *Proceedings of the Conference on Burnout Detection*, volume 38, 04 2021.

[3] Angelina R. Wilton, Katharine Sheffield, Quantia Wilkes, Sherry Chesak, Joel Pacyna, Richard Sharp, Paul E. Croarkin, Mohit Chauhan, Liselotte N. Dyrbye, William V. Bobo, and Arjun P. Athreya. The burnout prediction using wearable and artificial intelligence (brownie) study: a decentralized digital health protocol to predict burnout in registered nurses. volume 23, February 2024.

[4] Ziang Tang, Zachary King, Alicia Segovia, Han Yu, Gia Braddock, Asami Ito, Ryota Sakamoto, Motomu Shimaoka, and Akane Sano. Burnout prediction and analysis in shift workers: Counterfactual explanation approach. pages 1–4, 10 2023.

[5] Daniel McDuff, Isaac Galatzer-Levy,Seamus Thomson,Andrew Barakat,Conor Heneghan, Samy Abdel-Ghaffar, Jake Sunshine,Ming Zher-Poh,Lindsey Sunden,John Hernandez, Allen Jiang,Xin Liu,Ari Winbush,Benjamin W. Nelson, Nicholas B. Allen. Evidence of Differences in Diurnal Electrodermal, Temperature, and Heart Rate Patterns by Mental Health Status in Free-Living Data. August 2024.

[6] Samantha Ammons Erin L Kelly Phyllis Moen, Jack Lam. Time Work by Overworked Professionals: Strategies in Response to the Stress of Higher Status. April 2013.

[7] Wei Wang Yaofei Xie, Mengdi Ma. Trajectories of depressive symptoms and their predictors in Chinese older population: Growth Mixture model. June 2023.

[8] África Martos Martínez José Jesús Gázquez Linares María del Carmen Pérez-Fuentes, María del Mar Molero Jurado. Analysis of the Risk and Protective Roles of Work-Related and Individual Variables in Burnout Syndrome in Nurses. October 2019.