Leveraging LLM-Generated Embeddings for Topic Modeling and Survival Analysis of Obstructive Sleep Apnea Patients

Himanshu Yadav and Destinee Morrow

July 26,2024

Abstract

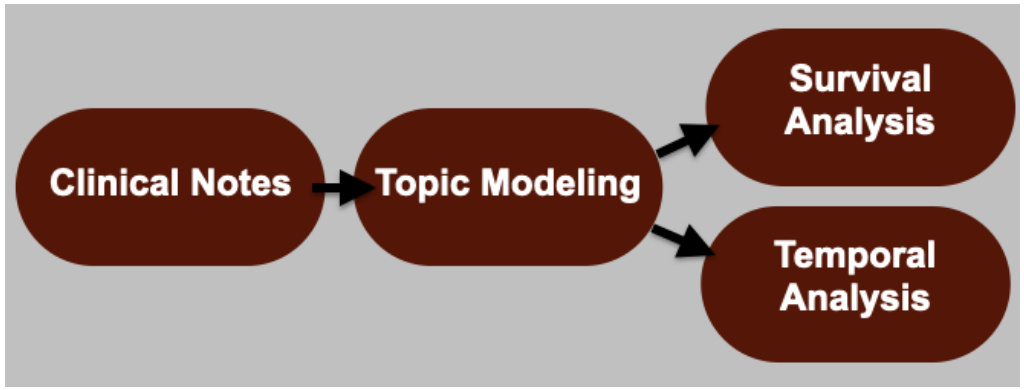
- This study leverages the MIMIC-IV database to analyze clinical narratives from the discharge notes of patients diagnosed with Obstructive Sleep Apnea (OSA) using BERTopic for topic modeling and survival analysis.
- Identified key themes such as "Symptoms and Diagnosis," "Treatment and Management," and "Follow-up and Monitoring."
- Temporal analysis revealed how these themes evolved over time.
- Kaplan-Meier survival analysis examined the impact of various covariates on patient outcomes.
- Novel insights into OSA management and significant factors affecting patient survival were revealed.

Background Info

- **Obstructive Sleep Apnea:**
  - Prevalent sleep disorder affecting health by causing reduced airflow during sleep, leading to various health problems.
  - Associated with hypertension, cardiovascular disease, and higher healthcare burdens.
- **MIMIC-IV Database:**
  - Comprehensive database containing patient data from clinical care units.
  - Includes diverse data types such as clinical notes, vitals, lab events, and medications.
- **BERTopic:**
  - Utilizes transformer-based language models to extract themes from text-based data.
- **GatorTron:**
  - Specialized language model trained on clinical texts to generate medical terminology and identify OSA management themes.

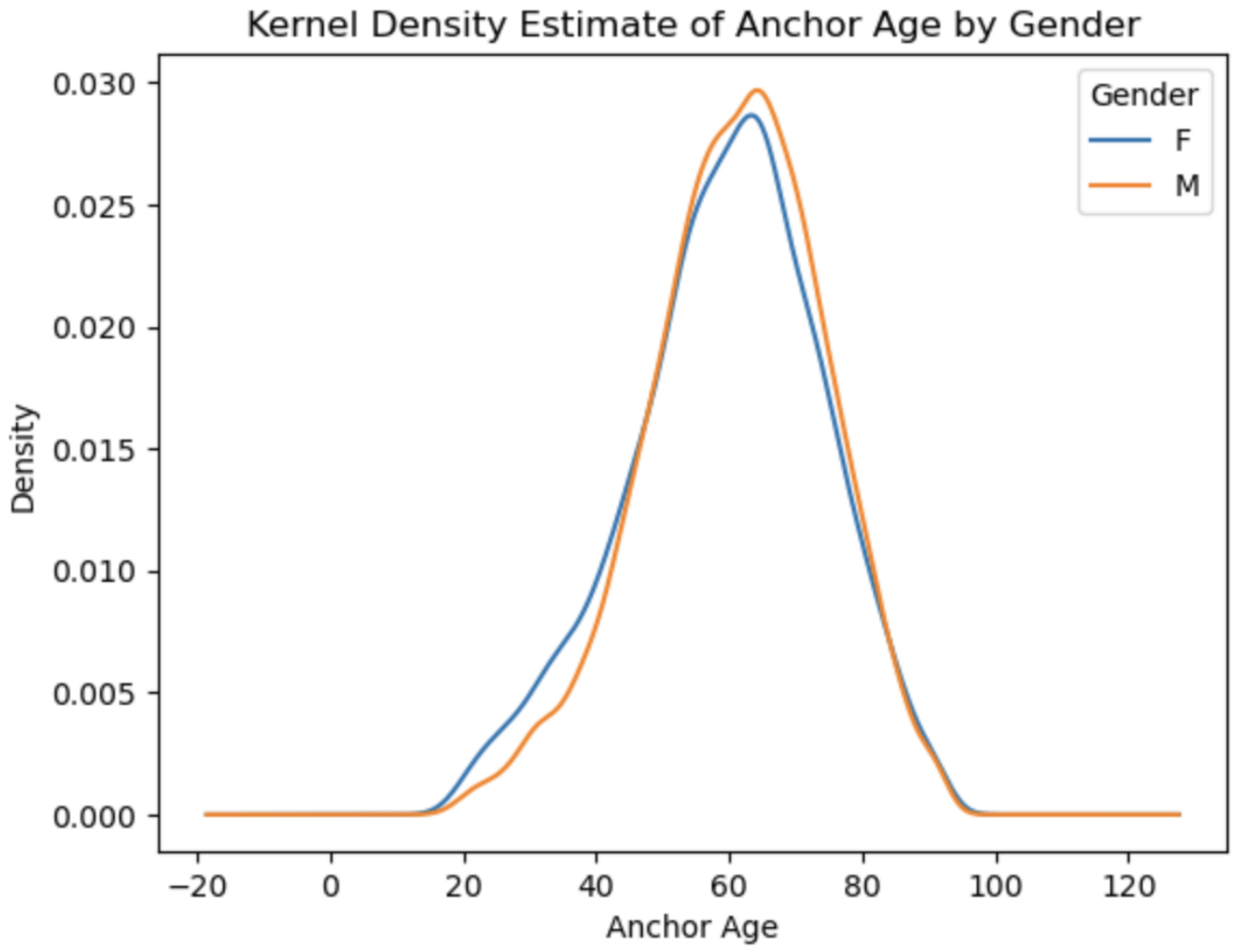
Research Hypothesis

- Advanced topic modeling using LLM, combined with survival analysis, can uncover evolving clinical themes and significant factors affecting patient outcomes in OSA management.



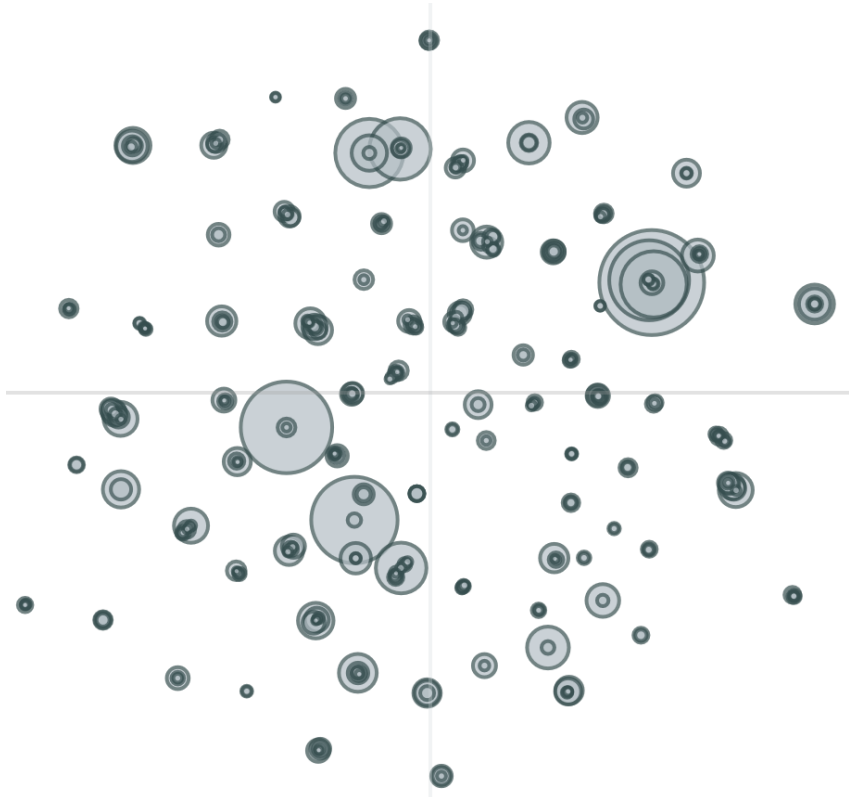
OSA Patient Demographics

- Total Number of Patients with OSA: 14,915
- Male Patients: 8,918
- Female Patients: 5,997



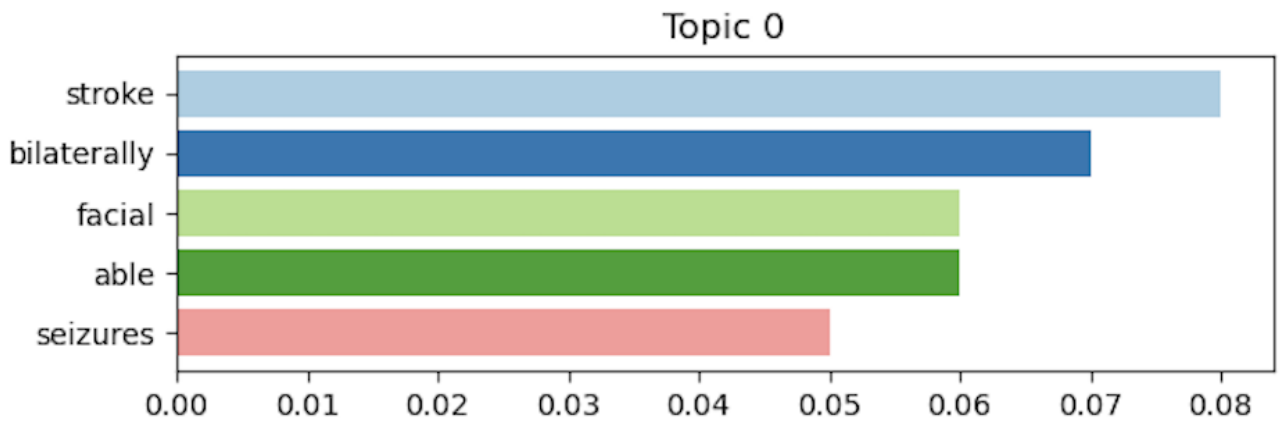
Topic Modeling

- Initial Analysis: Processed 51,865 clinical notes, identifying 350 topics.
- Refinement: Applied hierarchical clustering to merge overlapping topics, resulting in 166 distinct categories.

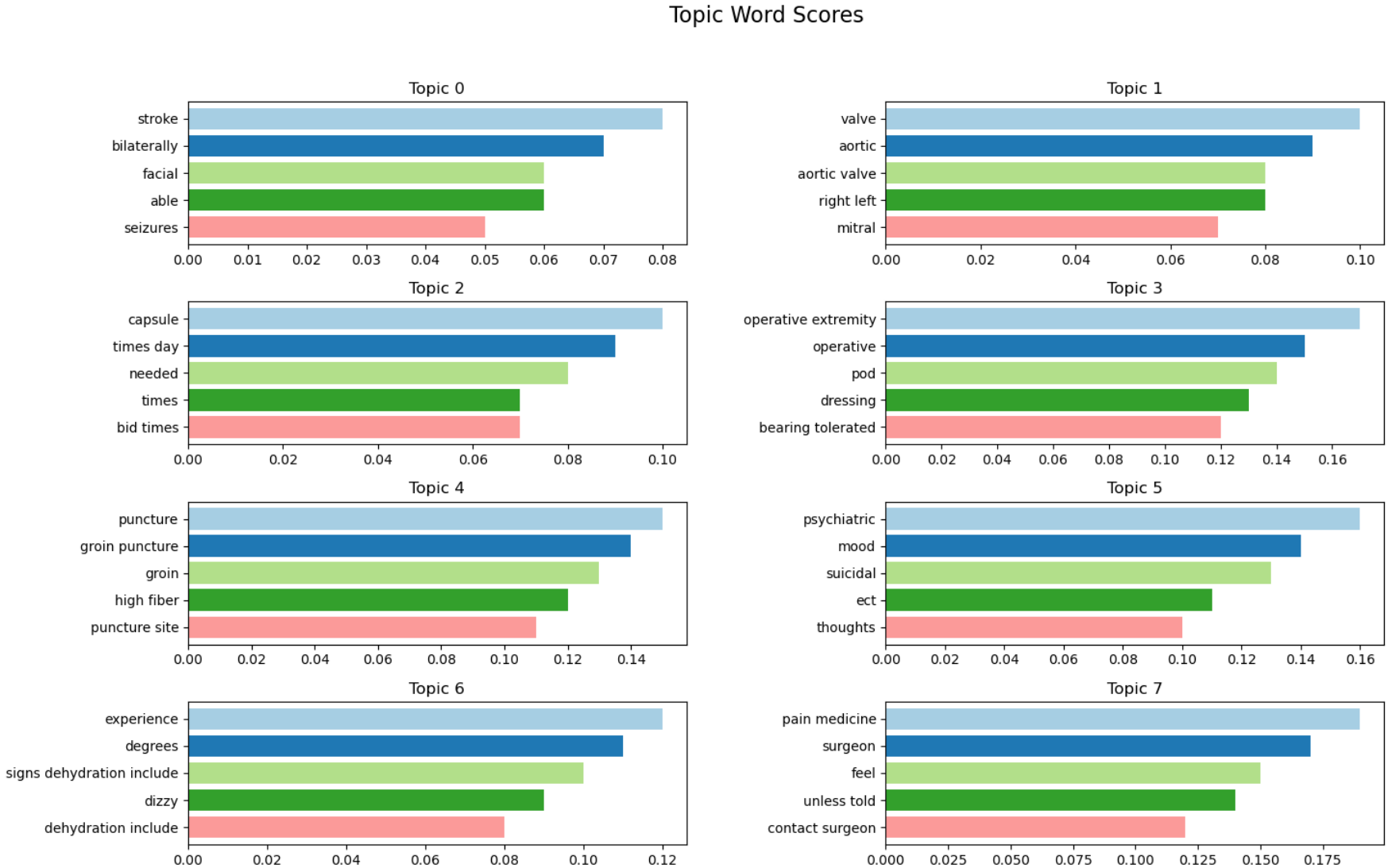


Themes in Clinical Notes

- Following bar plot represent the Topic 0, with the most significant words and their scores depicted to signify their importance within the topic.

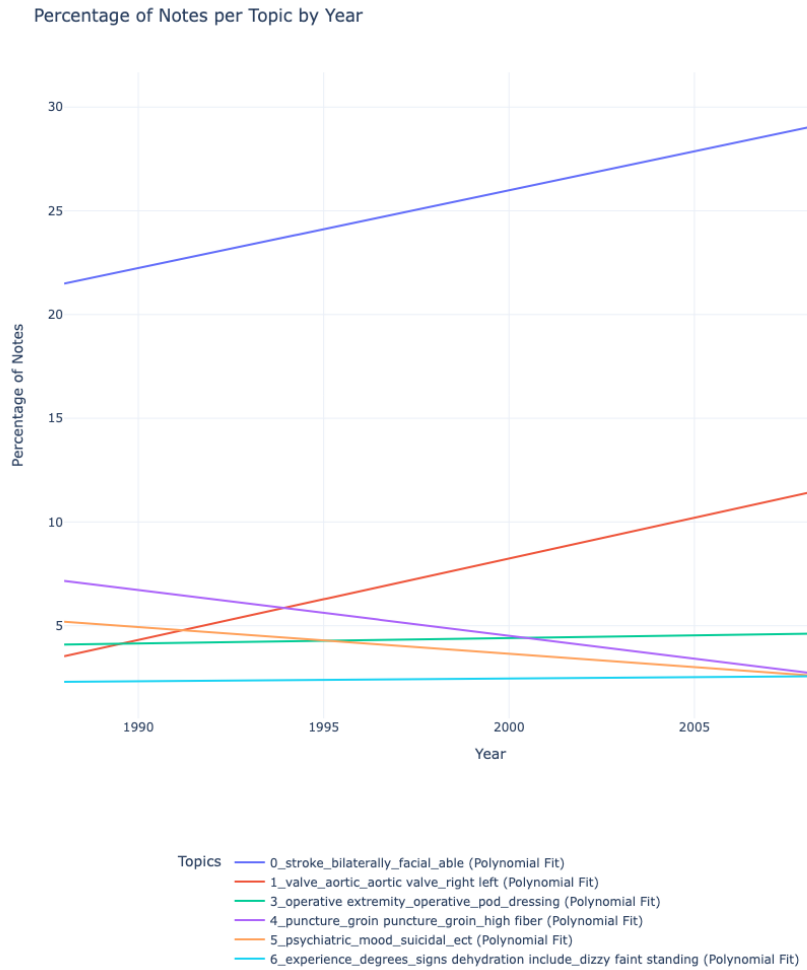


Visualization of the Top 8 Topics



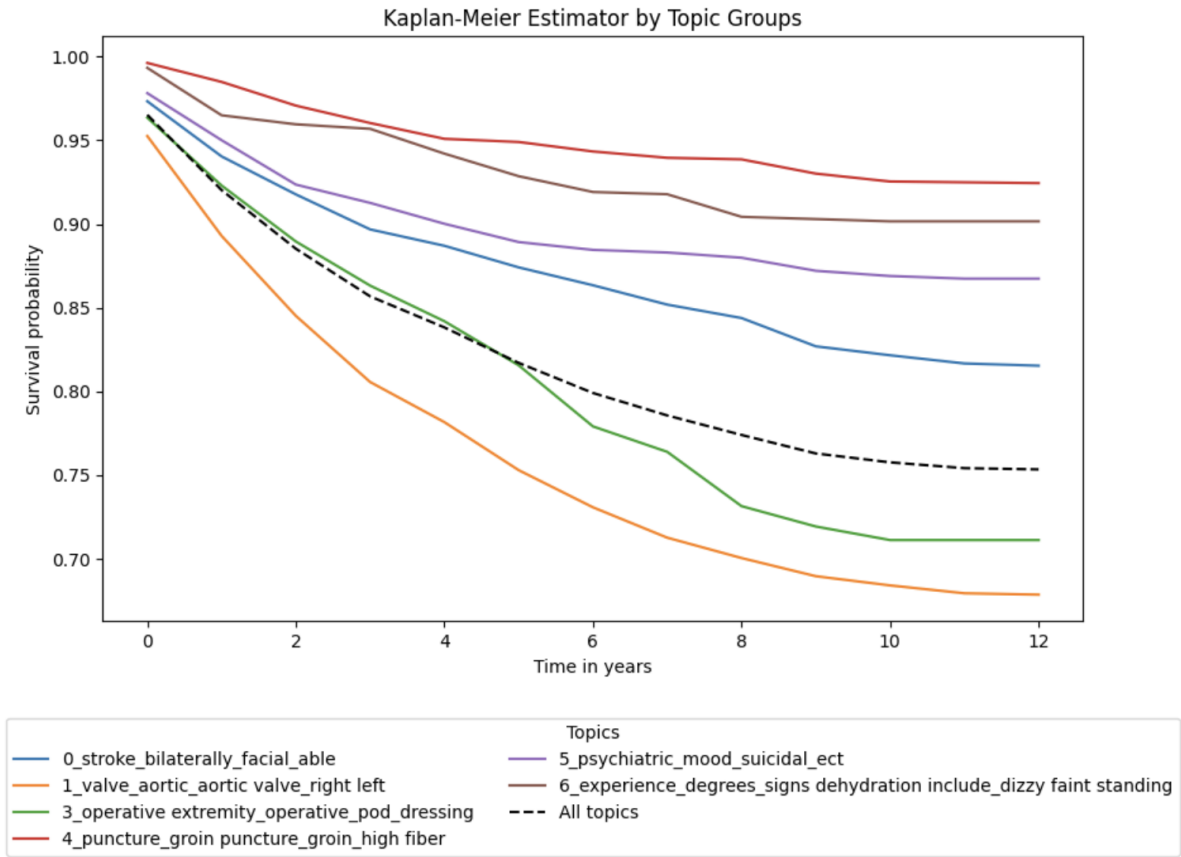
Temporal Analysis

- Temporal distribution of clinical notes across identified topics over time examining the diagnosed dates of OSA patients from 1987 to 2012.



Survival Analysis

- Kaplan-Meier survival curves for six topic groups derived from OSA patient discharge notes.
- Main takeaways indicate varying survival probabilities over 12 years.



Conclusion and Future Work

- Effectively identified crucial clinical themes and factors affecting patient outcomes in OSA management.
- Patients exhibit differing impacts of clinical themes on survival rates, guiding patient management strategies.
- Future Work: Refine topic modeling and survival analysis techniques, develop tailored treatment strategies, and translate identified themes to healthcare providers.

Topic 3

- Operative Extremity: This indicates that the surgeries are being performed on the limbs (extremities), which is common in orthopedic procedures.
- Weight Bearing Tolerated: This term is frequently used in orthopedic recovery, particularly in joint surgeries, to describe how much weight the patient can safely put on the operated limb.
- Mobilize: Encouraging early movement is typical in joint surgeries to promote recovery and prevent stiffness.
- Pod (Postoperative Day): Tracking recovery by days post-surgery is a standard practice in hospital settings for monitoring the patient's progress.