Pathological Cases Analysis at Hargeisa Group Hospital: A Descriptive Study

Dr. Omer Mohamed Bihi, Dr. Tex Li-Hsing Chi

2024-06-21

Table of contents

1 Abstract

This study presents a descriptive analysis of 110 pathological cases at Hargeisa Group Hospital (HGH) in Somaliland. The analysis covers surgical biopsies, fine-needle aspiration cytology (FNAC), cytology, and hematology cases. Key findings include a high prevalence of tuberculosis in lymph node specimens and anemia cases. The study also highlights the need for enhanced diagnostic capabilities and improved data management systems.

2 Introduction

Pathological analysis plays a crucial role in diagnosing and managing various diseases. In resource-limited settings like Somaliland, understanding the spectrum of pathological cases can inform healthcare planning and resource allocation. This study aims to provide insights into the types and frequencies of pathological cases seen at Hargeisa Group Hospital.

3 Methods

3.1 Data Collection

Data were extracted in Microsoft Access database and collected from the pathology department of Hargeisa Group Hospital between August 2023 and June 2024. The dataset includes 110 cases across four categories: surgical biopsies, FNAC, cytology, and hematology.

3.2 Data Interpretation with AI-powered approach

The pathology reports, initially recorded as free-text entries, were processed using natural language processing (NLP) techniques to extract non-structured information. We employed the Claude 3.5 Sonnet large language model (LLM), developed by Anthropic, to interpret and categorize the unstructured text data. This AI (artificial intelligent) powered approach allowed for efficient and consistent extraction of key diagnostic information, specimen types, and patient demographics from the English language used in the reports. The LLM was prompted to identify and classify diagnoses, categorize specimens, and extract relevant clinical details. To ensure accuracy, a two-step validation process was implemented: first, the AI-generated interpretations were reviewed by a trained medical data analyst, and then all cases was selected for review by an experienced surgeon doctor. This process allowed us to leverage the efficiency of AI while maintaining high standards of medical accuracy. The structured data output from this NLP pipeline was then compiled into standardized data tables, facilitating subsequent statistical analysis using R (version 4.4.0). This methodology enabled us to transform complex narrative reports into quantifiable data points, ensuring a comprehensive and reliable basis for our descriptive statistical analysis while minimizing human error in data extraction and classification.

3.3 Data Analysis

Data were managed by de-identification, cleaning, missing data management, and manual curation mentioned in "Data Interpretation with AI-powered approach". Descriptive statistics were calculated using R (version 4.4.0), copiloted with Claude 3 Opus LLM. Data were analyzed for frequency distributions, age ranges, and gender distributions.

4 Results

4.1 Surgical Biopsies

Table 1: Distribution of Surgical Biopsy Cases

Category	Count	Percentage (%)
Soft tissue lesions	7	29.2
Gynecological specimens	6	25.0
Gastrointestinal tract	3	12.5
Breast	2	8.3
Prostate	2	8.3
Other	4	16.7

Surgical Biopsy (24 cases): The surgical biopsies covered a diverse range of specimens, with soft tissue lesions (29.2%) and gynecological specimens (25.0%) being the most common. The age range was 2-80 years (mean: 41 years), with a slight female predominance (54.2%). - Soft tissue lesions (n=7, 29.2%): Including cases of epidermal inclusion cyst, intramuscular myxoma, lobular capillary hemangioma, and carcinoma in situ (origin of tissue site: not in record). - Gynecological specimens (n=6, 25.0%): Including cases of ruptured ectopic pregnancy, gestational choriocarcinoma, partial hydatidiform mole, high-grade serous carcinoma of the ovary, and leiomyoma. - Gastrointestinal tract (n=3, 12.5%): Two cases of inflammatory bowel disease (Crohn's disease) and one case of acute on chronic cholecystitis. -Breast (n=2, 8.3%): One case of benign phyllodes tumor and one case of pseudoangiomatous stromal hyperplasia. - Prostate (n=2, 8.3%): Both cases were diagnosed as benign nodular prostatic hyperplasia.

4.2 FNAC

```
library(knitr)
library(kableExtra)

fnac_data <- data.frame(
  Category = c("Lymph nodes", "Soft tissue lesions", "Breast lesions", "Thyroid", "Salivar
  Count = c(14, 13, 6, 5, 3, 5),
  Percentage = c(30.4, 28.3, 13.0, 10.9, 6.5, 10.9)
)

kable(fnac_data,
      col.names = c("Category", "Count", "Percentage (%)"),
      align = c("l", "c", "c"),
      digits = 1) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                full_width = FALSE) %>%
  footnote(general = paste("FNAC (46 cases): FNAC was most commonly performed on lymph nod
                           sep = "\n"))
```

Table 2: Distribution of Fine Needle Aspiratio

Category	Count	Pe
Lymph nodes	14	