Czech Salivary Gland Database

Vojtěch Jelínek[1][a], Petr Brůha[1][b], David Kalfeřt[2][c] and Pavel Nový[1][d]
[1]*Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, Pilsen, Czech Republic*
[2]*Department of Otorhinolaryngology and Head and Neck Surgery, 1st Faculty of Medicine, University Hospital Motol, Charles University, V Uvalu 84, Prague 150 06, Czech Republic*
vjelinek@students.zcu.cz, {pbruha, novyp}@kiv.zcu.cz, david.kalfert@fnmotol.cz

Keywords: Salivary Gland Cancer, Clinical Database, Patient Data Management, Data Visualization, Kaplan-Meier Analysis.

Abstract: Salivary gland tumors require comprehensive data collection and analysis to support clinical decision-making, yet existing databases need more focus on specific tumor-related data and visualization tools. This absence hinders oncologists' ability to track patient outcomes effectively and identify potential prognostic indicators. To address this, we developed the Czech Salivary Gland Database (CSGDB), a specialized clinical application designed to manage patient data and provide visual analytics. The database includes secure and anonymized data handling alongside Kaplan-Meier survival analysis for outcome visualization. Deployed at the University Hospital in Motol, CSGDB empowers healthcare professionals with enhanced tools for tracking and analyzing patient progress, ultimately contributing valuable data and insights to the field of head and neck oncology.

1 INTRODUCTION

Head and neck oncology represents a challenging field within medical research and clinical practice, given the diverse types of tumors, unique anatomical features, and complex treatment paths involved. Among these, salivary gland tumors are relatively rare yet require specialized care due to their heterogeneous nature, potential malignancy, and impact on critical physiological functions. Currently, effective diagnosis, treatment, and monitoring of salivary gland tumors rely on a combination of clinical expertise and robust, accurate data. However, existing medical databases and general-purpose electronic health record (EHR) systems must be better suited for managing the detailed, specific data required for this patient population.

The challenges associated with these tumors include tracking long-term outcomes, analyzing treatment effectiveness, and identifying potential prognostic markers—tasks that demand specialized databases with functionalities tailored to salivary gland oncology. Conventional data systems, such as general cancer registries, often lack fields for the specific parameters relevant to these tumors, limiting their usefulness for clinicians and researchers focused on this area. Furthermore, these databases frequently do not incorporate tools for survival analysis, data visualization, or detailed cohort tracking, which are essential for understanding disease progression and informing treatment decisions.

To address these limitations, we developed the Czech Salivary Gland Database (CSGDB) (Kalfeřt et al., 2024), a clinical application to support storing, managing, and analysis of patient data specific to salivary gland tumors. CSGDB provides a secure, specialized platform for capturing detailed clinical and demographic data, tracking treatment outcomes, and visualizing survival trends through integrated Kaplan-Meier analysis. The design of CSGDB not only addresses the data storage needs but also emphasizes user accessibility, ensuring that clinicians can efficiently utilize the system within their workflows.

This paper presents an overview of CSGDB's design, architecture, and functionalities. We also discuss the implementation and deployment of the system at the University Hospital in Motol, highlighting its impact on clinical workflow and its potential contributions to oncology research. CSGDB seeks to enhance patient care and research capabilities by offering a focused, clinically relevant database, laying

[a] https://orcid.org/0009-0009-8021-4204
[b] https://orcid.org/0000-0003-4031-8243
[c] https://orcid.org/0000-0003-4369-4113
[d] https://orcid.org/0000-0001-9945-1351

Jelínek, V., Brůha, P., Kalfeřt, D. and Nový, P.
Czech Salivary Gland Database.
DOI: 10.5220/0013254400003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 705-711
ISBN: 978-989-758-731-3; ISSN: 2184-4305

a foundation for future studies and improved clinical outcomes in head and neck oncology.

2 STATE OF THE ART

As described in the introduction, no specialized head and neck cancer database exists in the Czech Republic. From our point of view, it should be a comprehensive database focused on head and neck cancers, collecting complete clinicopathological data and subsequently used both for scientific purposes and in everyday practice. Managing a local database reflects the conditions and possibilities at that particular workplace (region, state); thus, the collected data's value should be higher. In addition to collecting and organizing data, such a database should contain tools for analyzing the various parameters being monitored. Other features include the ability to update the data for the monitored patients. Currently, most databases are based more on retrospective data collection, where patient data is entered on a specific date (most often the date the patient is entered into the database) without subsequent updates. Another advantage of a local database may be that patients with, e.g., a specific histopathological type of tumor can be easily retrieved, followed by tracing the actual tumor tissue in the pathology archives for further, e.g., molecular analyses.

The absence of such a database in the Czech Republic is not only on a national scale but also on the level of local databases of individual clinical departments.

The national database for head and neck cancer established by The Danish Head and Neck Cancer Study Group, DAHANCA, see (Overgaard et al., 2016), which is unique in the world and is designed for all head and neck cancers, is the inspiration for the creation of a local CSGDB with aspirations for future expansion. Its development from a local to a national database is also inspiring.

In addition to the DAHANCA database, other more general databases also contain datasets of head and neck cancers. Two examples from several such databases are:

- The International Collaboration on Cancer Reporting (ICCR) produces common, internationally validated and evidence-based pathology datasets for cancer reporting for use throughout the world through broad collaboration between Pathology Colleges, Societies and major cancer organizations internationally, spec. Head and Neck, see (Thompson et al., 2024), (Seethala et al., 2018);
- The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI), https://seer.cancer.gov/; see (Hankey et al., 1999), (Friedman and Negoita, 2024).

Common to these databases is that most are primarily focused on histopathological parameters of tumours and are therefore more intended for pathologists, i.e. complete clinicopathological data are missing.

Since we are currently dealing with large salivary gland tumors, we designed and implemented a CSGDB, which is being created within a clinical department with data collection from other clinical departments using a web form (https://slinnezlazy.cz/).

For an idea of the population and incidence of head and neck cancers in relation to the cited DANANCA database as an inspirational model, we present statistical data for the Czech Republic and Denmark for the period 2022; the data source is the International Agency for Research on Cancer, see (Ferlay et al., 2024), (Bray et al., 2024), Table 1.

Table 1: Overview table of newly diagnosed head and neck cancers in the Czech Republic and Denmark in 2022.

CZECHIA - 2022				
Population	10 736 782			
Number of new Cancer	65 676			
Cancer	**Number**	**(%)**	**Rank**	**Cum. risk**
Lip, oral, cavity	955	1.45	15	0.54
Oropharynx	582	0.89	21	0.36
Larynx	465	0.71	23	0.28
Salivary glands	151	0.23	27	0.07
Hypopharynx	150	0.23	28	0.10
Nasopharynx	51	0.08	32	0.03
Head & Neck	**2 354**	**3.58**		

Available from: https://gco.iarc.who.int/media/globocan/factsheets/populations/203-czechia-fact-sheet.pdf

DENMARK - 2022				
Population	5 834 952			
Number of new Cancer	48 840			
Cancer	**Number**	**(%)**	**Rank**	**Cum. risk**
Lip, oral, cavity	449	0.92	19	0.47
Oropharynx	450	0.92	18	0.52
Larynx	235	0.48	23	0.25
Salivary glands	82	0.17	28	0.08
Hypopharynx	120	0.25	27	0.13
Nasopharynx	21	0.04	32	0.03
Head & Neck	**1 357**	**2.78**		

Available from: https://gco.iarc.who.int/media/globocan/factsheets/populations/208-denmark-fact-sheet.pdf

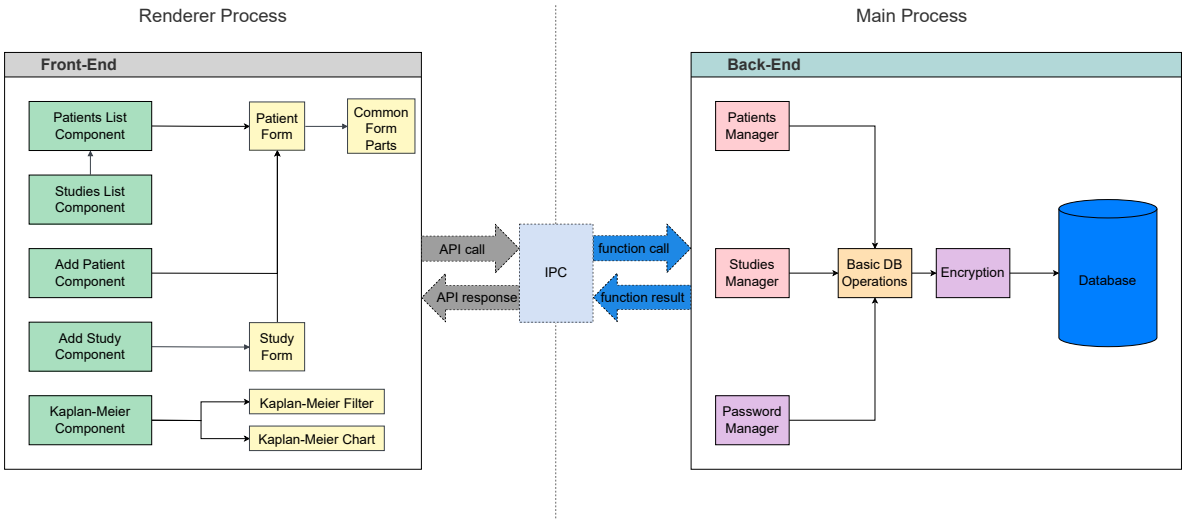

Figure 1: High-level Architecture of the Czech Salivary Gland Database (CSGDB).

3 SYSTEM DESIGN AND ARCHITECTURE

3.1 Overall Design Goals

The Czech Salivary Gland Database (CSGDB) design focuses on three main objectives: clinical usability, data security, and scalability. These priorities ensure the system functions effectively in a fast-paced clinical environment while adhering to strict data protection standards. CSGDB features a modular architecture that allows for future expansions, enabling the implementation of additional functionalities or system integrations without disrupting core operations.

The primary design objectives for CSGDB are to facilitate comprehensive data entry for patient records, support straightforward data retrieval and analysis, and ensure compliance with legal requirements for patient data security, such as GDPR standards. The system utilizes a modern technology stack that includes ElectronJS for cross-platform application development, ReactJS for creating a responsive and intuitive user interface, SQLite as the primary database for reliable data storage and portability, and NodeJS as the backend framework. This combination ensures that CSGDB is lightweight and high-performing, providing clinicians with an efficient and fast system. It is easily deployable on various operating systems used in healthcare facilities, further enhancing its usability.

The interface design enhances clinical usability by prioritizing ease of navigation and efficient data entry, allowing clinicians to input and access patient information quickly. The layout minimizes cognitive load, featuring customizable fields tailored to salivary gland tumors. These fields capture essential clinical and demographic details in an organized manner. This level of customization promotes consistency in data collection, which is beneficial for tracking patients over time and analyzing outcomes.

Data security is paramount in CSGDB, given the sensitive nature of patient records. The system incorporates encryption mechanisms, access control, and anonymization protocols to protect patient identities and ensure that only authorized personnel can access or modify data. This approach safeguards privacy and supports data sharing for research purposes, where anonymized patient data can drive insights while maintaining confidentiality.

Lastly, scalability is a key design consideration, allowing CSGDB to grow alongside advancements in clinical informatics and data management. The system's architecture and technology choices support efficiently handling an increasing volume of data. Future enhancements, such as machine learning models for predictive analysis, can be implemented with minimal changes to the existing infrastructure. This flexibility ensures that CSGDB remains a valuable resource for both clinical practice and research over time.

3.2 Architecture Overview

CSGDB is designed as a client-server (Oluwatosin, 2014) application (see Figure 1), providing a clear separation between data management on the server side and user interaction on the client side. The server

manages data storage and processing, while the client provides healthcare professionals with intuitive access to the system's functionalities through a user-friendly interface.

As illustrated in Figure 1, the CSGDB system architecture is built around a relational database back-end that handles both structured medical data (e.g., patient demographics, tumor classifications) and unstructured notes (e.g., clinician observations). This database forms the core of the application, ensuring reliable and secure data storage. An API layer facilitates communication between the database and client applications, providing seamless access to data.

We designed the client side to provide a responsive, cross-platform user experience that functions smoothly across different operating systems. The client-side application is structured around several key components to address the diverse needs of healthcare professionals. These include the patients list component, which displays a searchable and sortable list of patients; the studies list component, which organizes studies and their associated patient data for efficient navigation; the add patient component, allowing users to add new patient records seamlessly; the add study component, facilitating the creation of new studies with relevant patient associations; and the Kaplan-Meier component, enabling clinicians to generate survival curves for advanced data analysis.

On the server side, the back-end processes incoming requests, manages interactions with the database, and handles the application's core business logic. One of the most essential modules in the back-end is the encryption module, which is responsible for encrypting all the sensitive patient data.

Communication between the front-end and back-end is achieved with the Inter-Process Communication module. Its purpose is to process user requests from the front-end, then call appropriate functions on the back-end and provide a response with the result of the desired operation.

In addition to the core client-server architecture, the system's modular structure supports robust scalability. Future enhancements, such as adding predictive analytics or extending data processing capabilities, can be seamlessly integrated. The clear separation of responsibilities within the architecture ensures the secure handling of sensitive patient data.

3.3 Data Security and Anonymization

Data security within CSGDB is implemented through robust encryption, access control, and anonymization protocols. All sensitive patient data are encrypted upon entry and stored securely in the database, ensuring protection against unauthorized access. To further safeguard patient privacy, pseudonymization techniques are applied so that direct identifiers are separated from clinical information, minimizing the risk of re-identification. Access control is managed through two-factor authentication, allowing only authorized users to view or edit specific data. These measures collectively ensure a high data security and privacy standard, enabling secure handling of sensitive medical information in clinical and research settings.

4 KEY FUNCTIONALITIES OF CSGDB

CSGDB includes several key functionalities designed to streamline the data management process for clinicians and researchers, enabling better data utilization and visualization in clinical decision-making.

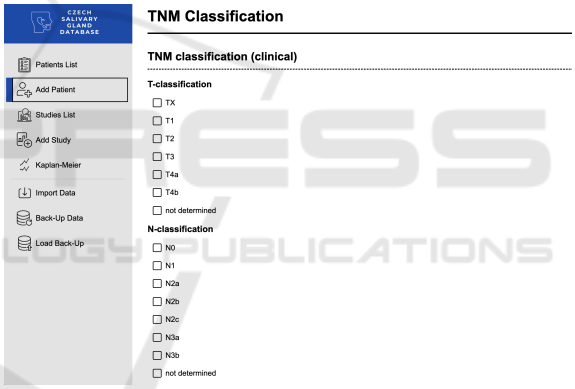

Figure 2: Screenshot of Czech Salivary Gland Database User Interface (Brierley et al., 2017).

4.1 User Interface and Experience

Our main goal for the CSGDB user interface was to make it straightforward for clinical staff to access. We devised a simple design that included two main components to accomplish this. The main menu is on the program window's left side, allowing users to access the application's core capabilities immediately. The corresponding view for the chosen menu option is displayed in the center of the application's windows (see Figure 2). This layout enables users to navigate between each functionality easily.

4.2 Updating Patient Records

Updating patient records is essential for medical professionals. Currently, most databases rely on retro-

spective data collection, meaning that patient information is recorded on a specific date—typically the date the patient is entered into the system—without any subsequent updates. Our application addresses this limitation by enabling the editing of patient records with a streamlined, user-friendly process.

The update begins when the user opens the patient's list component, displaying all patient records. The user can then locate the specific patient they wish to edit using search and filtering functionalities. Once the desired patient is identified, a single click on the edit button enables editing mode. In this mode, the user can modify the patient's details.

The interface provides two options upon completing the edits. If the user is satisfied with the changes, they can click the save button, which commits the updated data to the database. Alternatively, if the user decides to discard the changes, they can click the undo button, which reverts the record to its previous state.

4.3 Application Security

Because clinical data is one of the most sensitive pieces of information about anyone's life, it is necessary to secure it properly. Our application encrypts all sensitive information about patients with AES encryption (Daemen, 1999), (Rijmen and Daemen, 2001). To log in to the application, the user must provide his password simultaneously with the encryption key; otherwise, access to the application is not granted.

4.4 Data Filtration

The ability to filter patients based on defined factors is critical for effectively utilizing clinical datasets. Currently, the application supports filtration based on parameters such as the type of impacted salivary gland, the therapy administered, and the histopathological classification. The system's modular design facilitates the addition of new filtering criteria based on user requirements or emerging research needs.

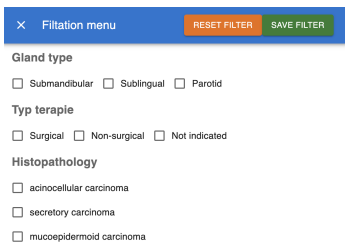

Figure 3: Filtration menu.

The filtration menu, accessible within the patient's list component and illustrated in Figure 3, allows users to configure and apply filters based on various criteria. Users can also reset all applied filters to their default state. These functionalities enhance the system's flexibility, supporting targeted data analysis, which proves especially valuable in clinical and research contexts.

4.5 Visualization and Analytics

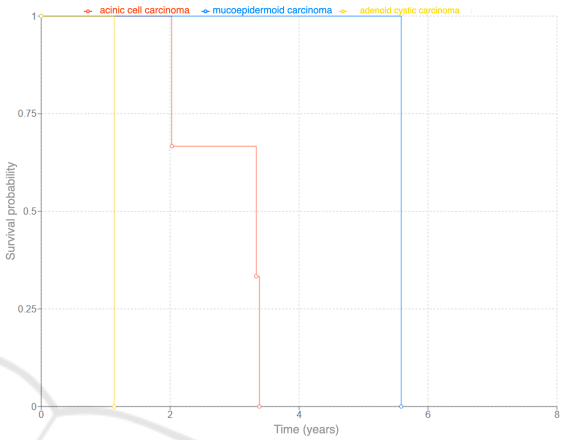

Figure 4: Generated Kaplan-Meier curve.

The application offers users a tool to generate Kaplan-Meir (Kaplan and Meier, 1958) curves. Currently, there are two possible curve variations. The first represents survival probability, while the second depicts recurrence likelihood.

Patients for the curve dataset are chosen based on their histopathological type. Multiple Kaplan-Meier curves can be displayed in the same chart, allowing quick visual data comparisons between patients with certain histology classifications. (Goel et al., 2010)

Figure 4 shows the Kaplan-Meier survival curves for patients across three histopathological groups listed at the top of the graph. The X-axis represents time in years, and the Y-axis indicates the probability of survival.

4.6 Data Back-Up and Import/Export

The CSGDB offers robust tools for managing data integrity and interoperability. Users can perform database back-ups, a critical feature given the extended timeframes often required for comprehensive data collection. These back-ups are securely stored and encrypted, ensuring that sensitive data remains protected even in unexpected damage to the user's machine. Each back-up file can be restored within the application by providing the correct password and encryption key, ensuring only authorized users can access the data. While the application does not cur-

rently include tools for advanced statistical analysis, it enables data exchange through an import/export feature. Users can export patient and study data into Excel files, allowing for detailed statistical processing in external software. This functionality ensures clinicians and researchers can leverage specialized analytical tools without compromising data security or usability.

Understanding the sensitive nature of patient data, the application provides an export option that anonymizes all personal patient data. This feature is particularly useful when medical professionals share their research data, ensuring patient privacy is always protected.

It is also possible to import the Excel file with data into CSGDB, which enables sharing of the collected data between application instances.

5 IMPLEMENTATION AND DEPLOYMENT

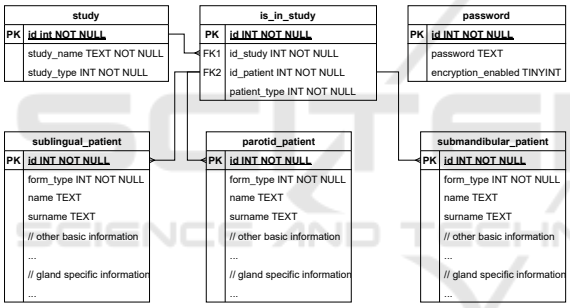

Figure 5: ERA Model of CSGDB.

The CSGDB prioritizes reliability and scalability, leveraging a modern tech stack to ensure secure, efficient, and compliant operations within a clinical environment.

5.1 Technology Stack and Data Model

The CSGDB is built on a modern technology stack that provides secure, scalable, and efficient data management. It employs an SQLite relational database for data storage, ensuring reliability and flexibility in handling clinical data. The backend is implemented using a Node.js-based API, while the front-end interface is developed with ReactJS, providing a responsive and user-friendly experience.

The underlying database schema is structured to meet the specific needs of salivary gland tumor data management. Figure 5 illustrates the Entity-Relationship-Attribute (ERA) model of the database, which comprises six tables, one for each gland that may be affected by a tumor. Each study can include n patients, and each patient can participate in n studies, reflecting the complex relationships inherent in clinical datasets.

Additionally, the database includes a `password` table to store hashed passwords and encryption-related metadata. Users can enable or disable encryption during the application's initial setup. This feature supports testing with fictitious data while maintaining security for actual clinical data.

5.2 Deployment Context

CSGDB was deployed at the University Hospital in Motol as a pilot program. This setting allowed for real-world testing and validation of the system's usability and performance.

6 CONCLUSIONS

The Czech Salivary Gland Database (CSGDB) is a significant contribution to clinical data management and oncological research. By addressing challenges such as secure storage, real-time updates, and data visualization, the platform provides a structured approach to managing data related to salivary gland tumors. The tailored design of CSGDB facilitates clinical decision-making and research, bridging a critical gap in the availability of domain-specific information systems.

The development of CSGDB involved integrating advanced technological components, including AES encryption for data security, modular architecture for extensibility, and tools for visualizing survival outcomes. Its deployment in a clinical setting demonstrates the system's applicability and reliability under real-world conditions. These achievements underscore the value of combining technological innovation with domain-specific requirements in developing medical information systems.

Future work will focus on extending the database to encompass additional rare tumor types, thereby increasing its applicability to a broader range of oncological studies. The planned integration of predictive analytics and large language models will further enhance the system's capacity to provide insights, contributing to advancements in personalized medicine. These directions align with the broader objectives of improving patient outcomes and facilitating high-quality research in head and neck oncology.

This work exemplifies the role of targeted databases in advancing clinical research and practice,

offering a model for future systems to address specific challenges in medical data management and analysis.

ACKNOWLEDGEMENTS

This work was supported by Grant No. SGS-2022-016 Advanced methods of data processing and analysis.

REFERENCES

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., and Jemal, A. (2024). Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3):229–263.

Brierley, J. D., Gospodarowicz, M. K., and Wittekind, C. (2017). *TNM Classification of Malignant Tumours*. John Wiley & Sons, Ltd, UICC, eighth edition.

Daemen, J. (1999). Aes proposal: Rijndael.

Ferlay, J., Ervik, M., Lam, F., Laversanne, M., Colombet, M., Mery, L., Piñeros, M., Znaor, A., Soerjomataram, I., and Bray, F. (2024). Global cancer observatory: Cancer today. Available from: https://gco.iarc.who.int/today, https://gco.iarc.who.int/media/globocan/factsheets/populations/208-denmark-fact-sheet.pdf, https://gco.iarc.who.int/media/globocan/factsheets/populations/203-czechia-fact-sheet.pdf.

Friedman, S. and Negoita, S. (2024). History of the surveillance, epidemiology, and end results (seer) program. *JNCI Monographs*, 2024(65):105–109.

Goel, M. K., Khanna, P., and Kishore, J. (2010). Understanding survival analysis: Kaplan-meier estimate. *International journal of Ayurveda research*, 1(4):274.

Hankey, B. F., Ries, L. A., and Edwards, B. E. (1999). The surveillance, epidemiology, and end results program: A national resource. *Cancer Epidemiology, Biomarkers & Prevention*, 8:1117–1121.

Kalfeřt, D., Jelínek, V., Nový, P., and Brůha, P. (2024). Czech salivary gland database - a computer software for managing a database of patients with malignant tumors of the major salivary glands. In *10th Czech-Slovak Congress, the 85th Congress of the Czech ENT Society of JEP, and the 70th Congress of the Slovak Society for ENT and Head and Neck Surgery*.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.

Oluwatosin, H. S. (2014). Client-server model. *IOSR Journal of Computer Engineering*, 16(1):67–71.

Overgaard, J., Jovanovic, A., Godballe, C., and Eriksen, J. (2016). The danish head and neck cancer database. *Clinical Epidemiology*, 8:491–496.

Rijmen, V. and Daemen, J. (2001). Advanced encryption standard. *Proceedings of federal information processing standards publications, national institute of standards and technology*, 19:22.

Seethala, R. R., Altemani, A., Ferris, R. L., Fonseca, I., Gnepp, D. R., Ha, P., Nagao, T., Skalova, A., Stenman, G., and Thompson, L. D. R. (2018). *Carcinomas of the Major Salivary Glands Histopathology Reporting Guide*. International Collaboration on Cancer Reporting, Sydney, Australia, 1st edition.

Thompson, L. D. R., Bishop, J. A., Hyrcza, M. D., Ihrler, S., Leivo, I., Nagao, T., Rupp, N. J., Skalova, A., Stenman, G., Vander Poorten, V., van Herpen, C., and Helliwell, T. (2024). *Carcinomas of the Major Salivary Glands Histopathology Reporting Guide*. International Collaboration on Cancer Reporting, Sydney, Australia, 2nd edition.