

# Assignment #1

## Exploring Cancer Data and Research Directions in Healthcare AI



<b>Course:</b>	Artificial Intelligence in Healthcare
<b>Level:</b>	MS (Graduate)
<b>Instructor:</b>	Dr. Naveed Anwer Butt
<b>Student Name:</b>	Hira Nawaz
<b>Roll No:</b>	002
<b>Institute:</b>	University of Gujrat, Hafiz Hayat Campus

## Introduction

Cancer remains one of the leading causes of death worldwide, representing a major public health challenge. Understanding the epidemiology of cancer and exploring how Artificial Intelligence (AI) can aid in its early detection is crucial for improving survival rates and treatment outcomes. This report explores global and national cancer statistics, focusing particularly on breast cancer in Pakistan. Furthermore, it highlights how AI-based approaches are being used for early detection, diagnosis, and data-driven insights using open-source datasets.

## Part A – Cancer Landscape Exploration

### 1.Global View

According to the World Health Organization (WHO) and GLOBOCAN 2022, more than 36 major cancer types are recognized globally. The top 10 most prevalent cancers worldwide include breast, lung, colorectal, prostate, stomach, liver, cervical, thyroid, bladder, and non-Hodgkin lymphoma. Breast cancer accounts for approximately 12.5% of all new cancer cases globally, making it the most diagnosed cancer worldwide.

#### Global Cancer Types (WHO / GLOBOCAN Classification)

#	Cancer Type / Site	ICD-10 Code(s)	Body Region / Description (in simple words)
1	Lip, Oral Cavity	C00–C06	Mouth and lips cancers
2	Salivary Glands	C07–C08	Glands producing saliva
3	Oropharynx	C09–C10	Throat (back of mouth)
4	Nasopharynx	C11	Upper throat behind the nose
5	Hypopharynx	C12–C13	Lower throat area
6	Oesophagus	C15	Food pipe connecting mouth to stomach
7	Stomach	C16	Stomach cancer
8	Colon	C18	Large intestine (first part)
9	Rectum	C19–C20	Lower part of the large intestine
10	Anus	C21	Anal canal
11	Liver (incl. intrahepatic bile ducts)	C22	Liver and bile duct cancers
12	Gallbladder	C23	Gallbladder cancer
13	Pancreas	C25	Pancreatic cancer
14	Larynx	C32	Voice box
15	Lung (incl. trachea)	C33–C34	Lungs and windpipe
16	Melanoma of Skin	C43	Skin cancer from pigment cells
17	Non-Melanoma Skin Cancer (NMSC)	C44	Common skin cancers (except melanoma)
18	Mesothelioma	C45	Cancer of the tissue lining organs (often lungs)
19	Kaposi Sarcoma	C46	Cancer linked with HIV infection

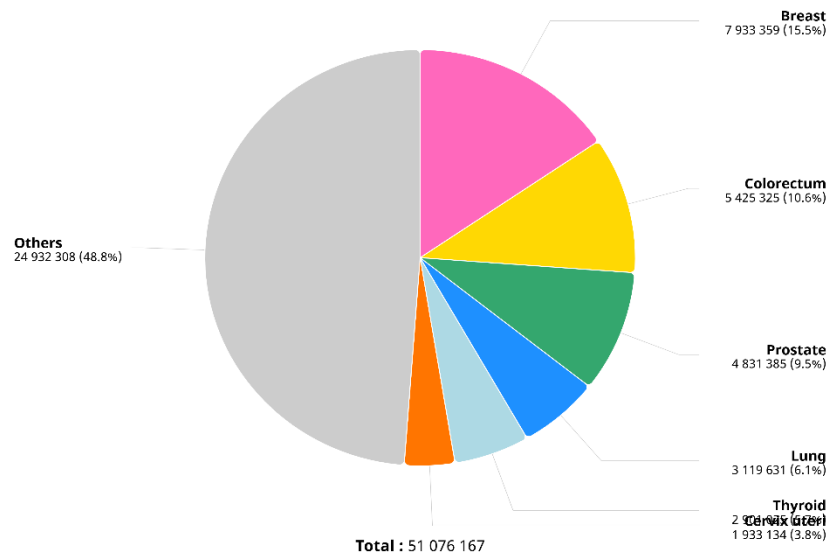
20	Female Breast	C50	Breast cancer
21	Vulva	C51	Outer part of female genital area
22	Vagina	C52	Inner part of female genital area
23	Cervix Uteri	C53	Lower part of uterus (cervix)
24	Corpus Uteri	C54	Body of uterus (endometrial cancer)
25	Ovary	C56	Ovarian cancer
26	Penis	C60	Male external genital organ
27	Prostate	C61	Prostate gland (male)
28	Testis	C62	Testicular cancer
29	Kidney	C64	Kidney cancer
30	Bladder	C67	Urinary bladder cancer
31	Brain, Central Nervous System	C70–C72	Brain and spinal cord cancers
32	Thyroid	C73	Thyroid gland cancer
33	Hodgkin Lymphoma	C81	Type of blood cancer affecting lymph nodes
34	Non-Hodgkin Lymphoma	C82–C86, C88	Other lymph node cancers
35	Multiple Myeloma	C90	Cancer of plasma cells in bone marrow
36	Leukaemia	C91–C95	Blood and bone marrow cancers
37	Other Specified Cancers	Various (C17, C24, etc.)	Includes rare or site-specific cancers
38	Unspecified Sites	C76–C80, C96–C97	Unknown or mixed location cancers
39	All Cancers (Combined)	C00–C97	All cancer sites grouped together

- **Total recognized cancer sites:** 39
- **ICD-10 codes** are used internationally to label diseases for data and research.
- **Grouped by:** body region (head & neck, digestive, respiratory, reproductive, etc.)
- **Special cases:**
  - *Bladder cancer* may include uncertain or in-situ tumors.
  - *Non-melanoma skin cancer (C44)* excludes basal cell carcinoma in incidence data.
  - *Kaposi sarcoma* is included in some HIV-related mortality statistics.

**Estimated number of prevalent cases (5-year), Both sexes, age [0-84], in 2022**

World

All cancers

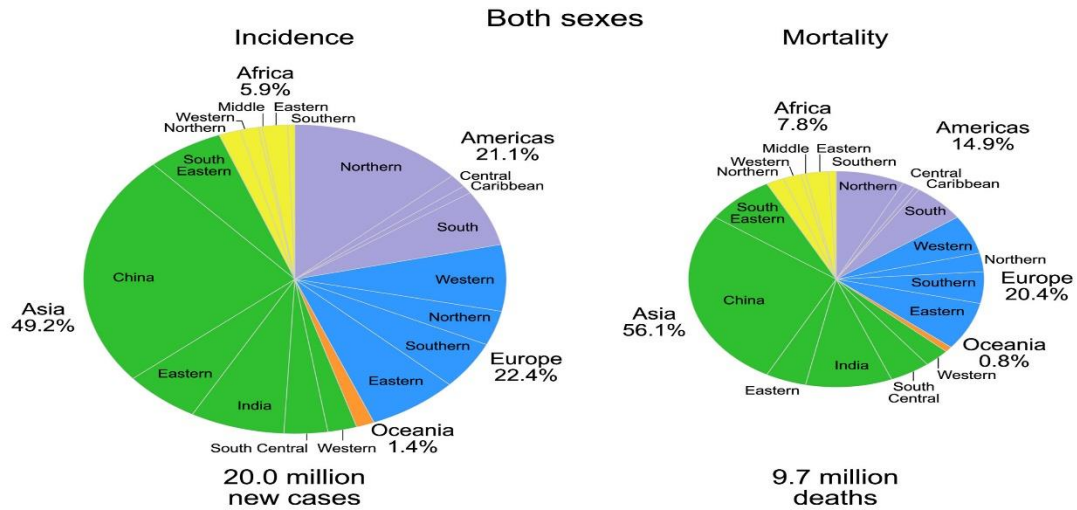


Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

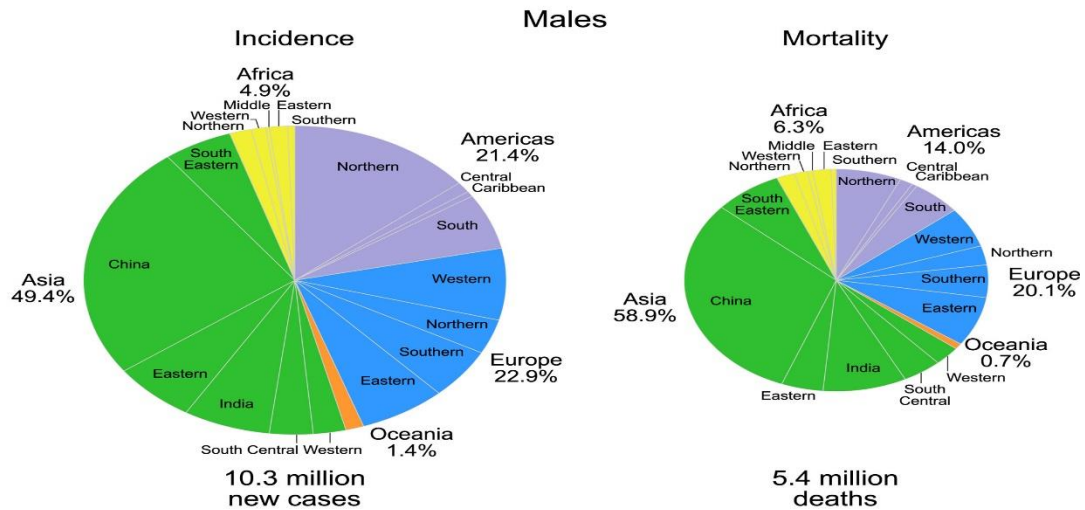


## Continents wise Prevalence of Cancer Types

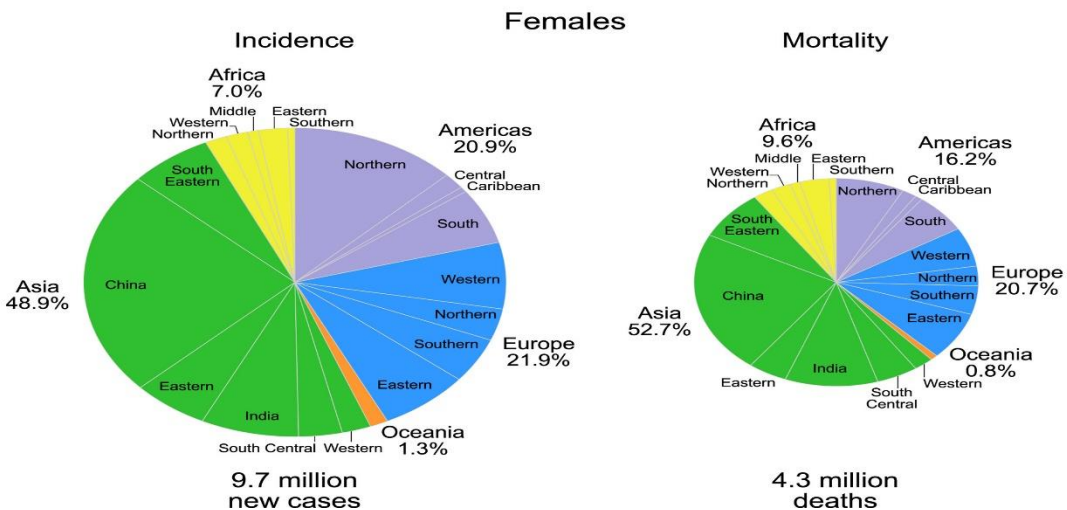
a)



b)



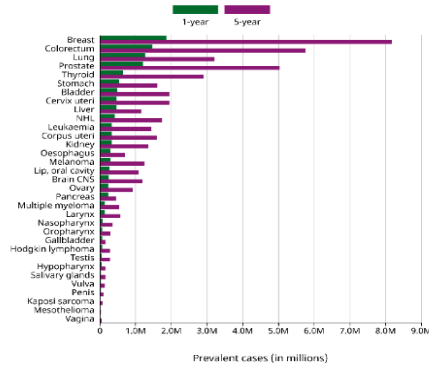
c)



## Prevalent Cases from Recent 1 to 5 years

Estimated number of prevalent cases, Both sexes, in 2022

World  
All cancers

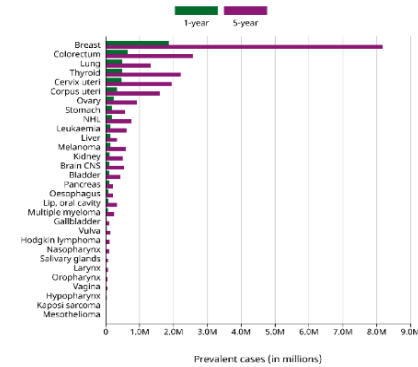


Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version: Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

Estimated number of prevalent cases, Females, in 2022

World  
All cancers

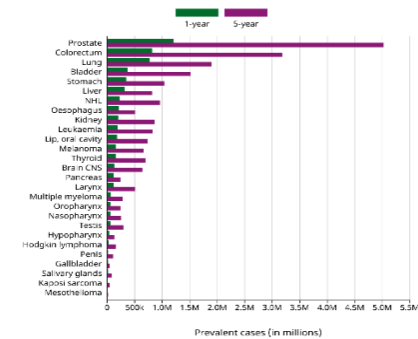


Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version: Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

Estimated number of prevalent cases, Males, in 2022

World  
All cancers



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version: Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

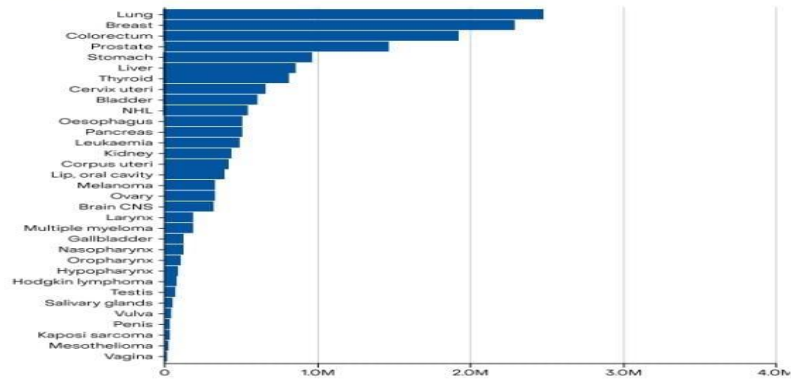
## Incidence

Cases  
**19 976 499**

ASR (World)  
**196.9**

### Cancer site ranking

#### | Incidence



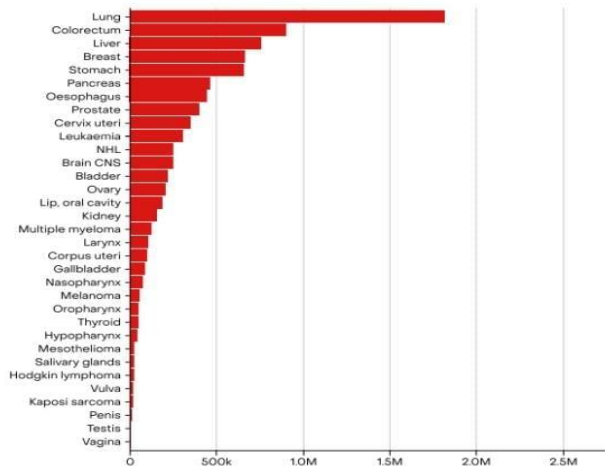
Number of new cases, both sexes, all ages

## Mortality

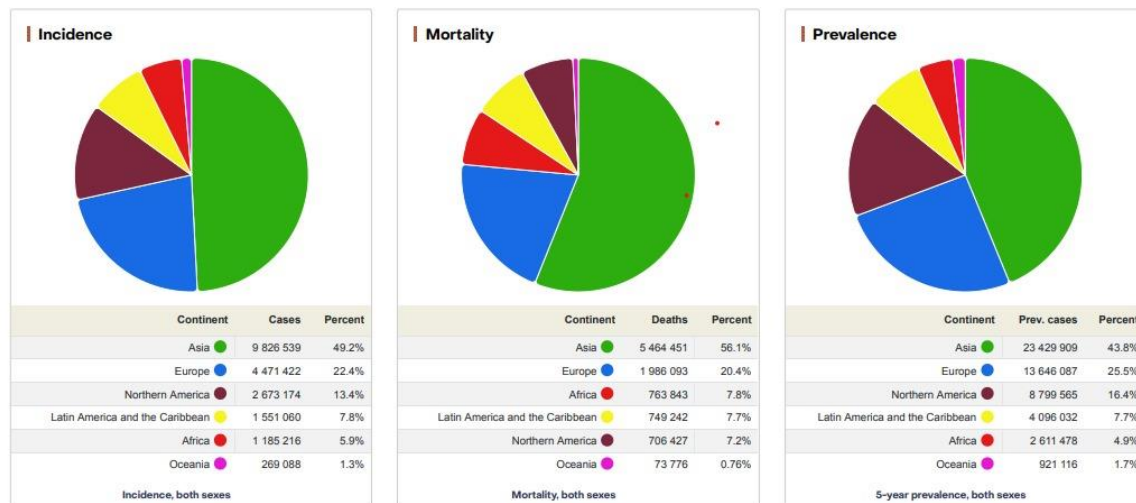
Deaths  
**9 743 832**

ASR (World)  
**91.7**

#### | Mortality



Number of deaths, both sexes, all ages



## Top 10 prevalent cancer types:

The top 10 cancer types in both sexes account for over 60% of newly diagnosed cancer cases and cancer deaths. Lung cancer is the most commonly diagnosed cancer worldwide (12.4% of the total cases), followed by cancers of the female breast (11.6%), colorectum (9.6%), prostate (7.3%), and stomach (4.9%). Lung cancer is also the leading cause of cancer death (18.7% of the total cancer deaths), followed by colorectal (9.3%), liver (7.8%), female breast (6.9%), and stomach (6.8%) cancers. In women, breast cancer is the most commonly diagnosed cancer and the leading cause of cancer death, followed by lung and colorectal cancer for both cancer cases and deaths; whereas lung cancer is most frequent cancer in men (both cases and deaths), followed by prostate and colorectal cancer for new cases and liver and colorectal cancer for deaths.

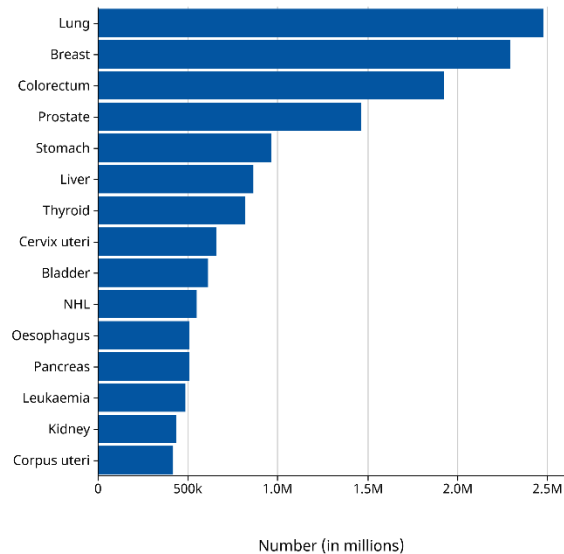


### Top 10 Cancer Types Globally (2022 Incidence)

Rank	Cancer Type	Approximate New Cases in 2022	% of All New Cancer Cases*
1	Lung (trachea, bronchus, lung)	~2,480,000	~12.4%
2	Female Breast	~2,300,000	~11.6%
3	Colorectum (colon + rectum)	~1,926,000	~9.6%
4	Prostate	~1,467,000	~7.3%
5	Stomach	~968,000	~4.9%
6	Liver & intrahepatic bile ducts	~865,000	~4.3%
7	Thyroid	~821,000	~4.1%
8	Cervix uteri (cervical cancer)	~661,000	~3.3%
9	Bladder	~613,000	~3.1%
10	Non-Hodgkin lymphoma	~553,000	~2.8%

#### Absolute numbers, Incidence, Both sexes, in 2022

World  
(Top 15 cancer sites)

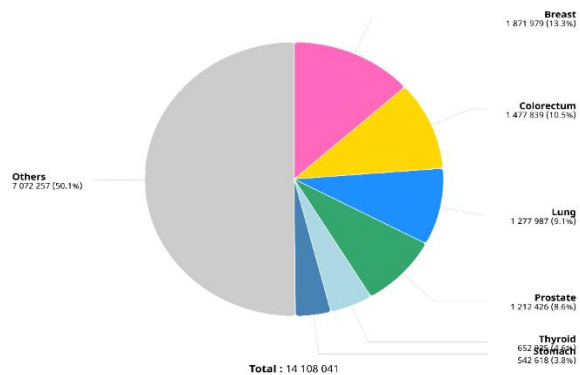


Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025



## From Recent 1, 3 and 5 years:

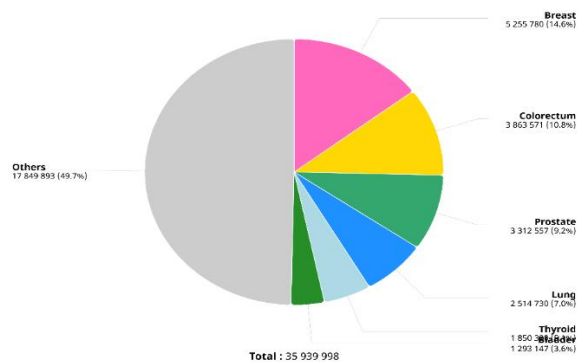
Estimated number of prevalent cases (1-year), Both sexes, in 2022  
World  
All cancers



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

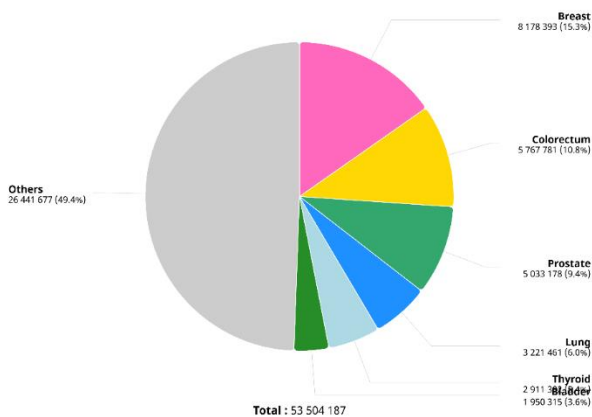
Estimated number of prevalent cases (3-year), Both sexes, in 2022  
World  
All cancers



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

Estimated number of prevalent cases (5-year), Both sexes, in 2022  
World  
All cancers



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

## 2.National View (Pakistan)

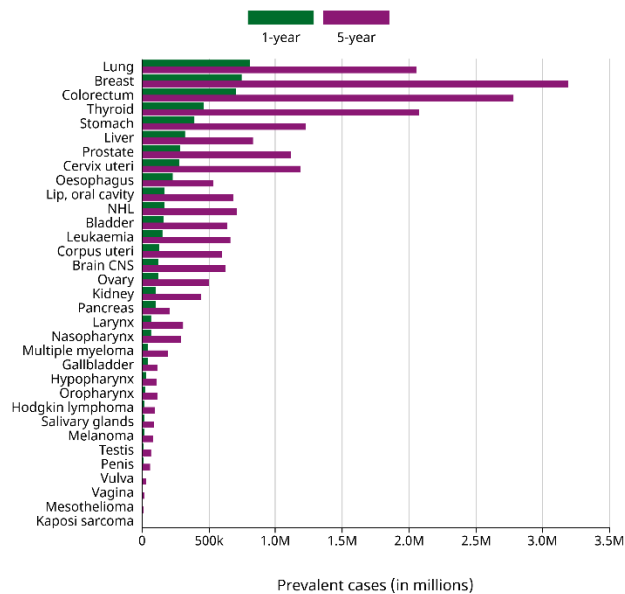
In Pakistan, **Shaukat Khanum Cancer Registry** and **GLOBOCAN 2022** report the following:

- Most prevalent cancers: **breast, oral cavity, lung, colorectal, lymphoma**
- **Gender trends:** Breast cancer dominates in females, oral and lung cancers dominate in males
- **Region trends:** Punjab has the highest incidence, followed by Sindh and Khyber Pakhtunkhwa

Estimated number of prevalent cases, Both sexes, in 2022

Asia

All cancers

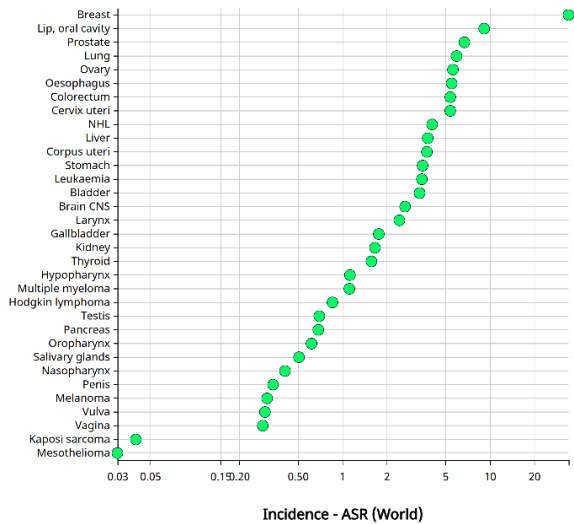


Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

Cancer site vs Incidence - ASR (World), Both sexes, in 2022

Pakistan

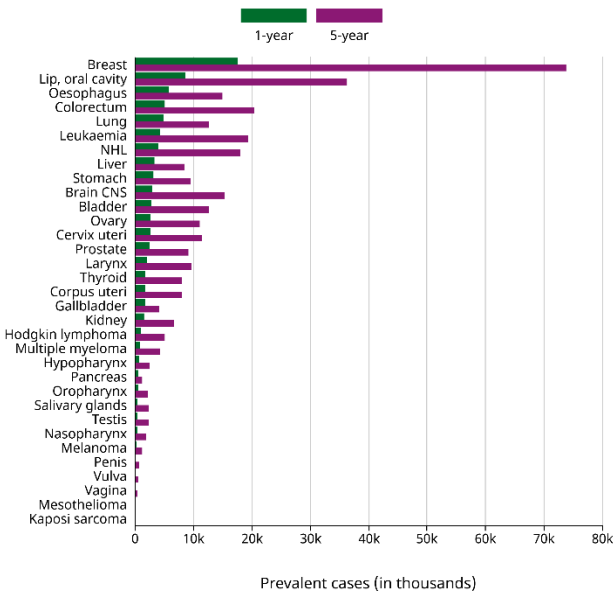


Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025



Estimated number of prevalent cases, Both sexes, in 2022

Pakistan  
All cancers

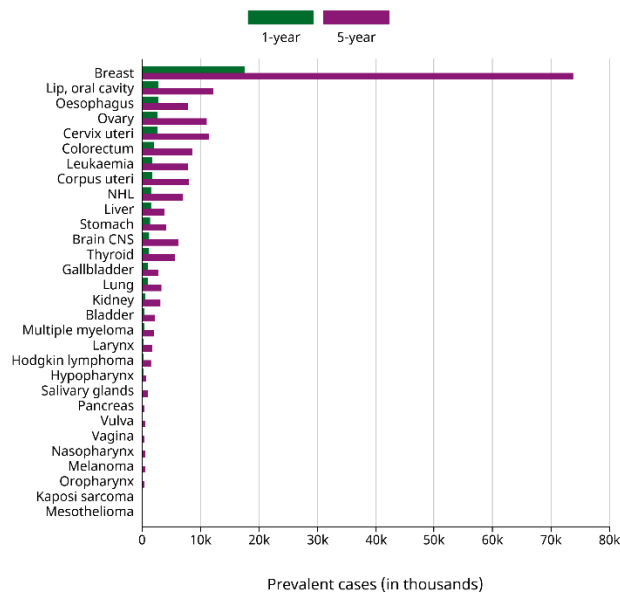


Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025



### Estimated number of prevalent cases, Females, in 2022

Pakistan  
All cancers

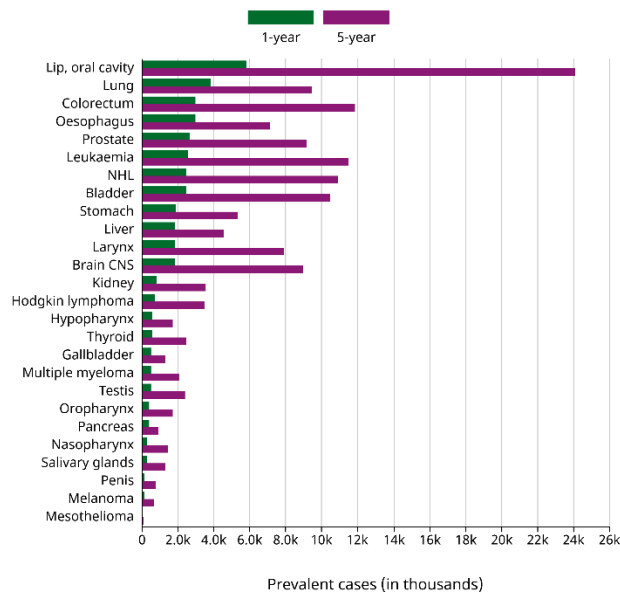


Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025



### Estimated number of prevalent cases, Males, in 2022

Pakistan  
All cancers



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

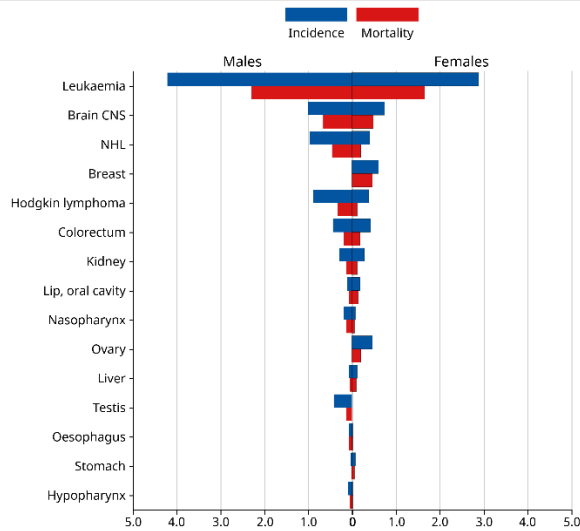


## According to incident ,mortality and Age

Age-Standardized Rate (World) per 100 000, Incidence and Mortality, Males and Females, age [0-24], in 2022

Pakistan

(Top 15 cancer sites)



ASR (World) per 100 000

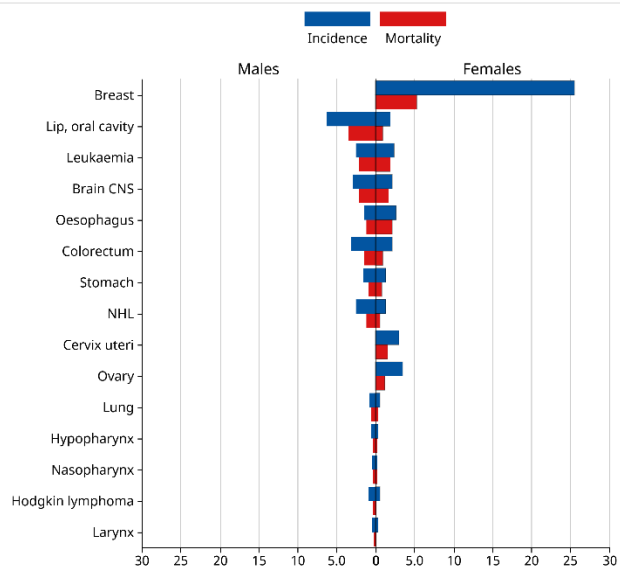
Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025



Age-Standardized Rate (World) per 100 000, Incidence and Mortality, Males and Females, age [25-39], in 2022

Pakistan

(Top 15 cancer sites)

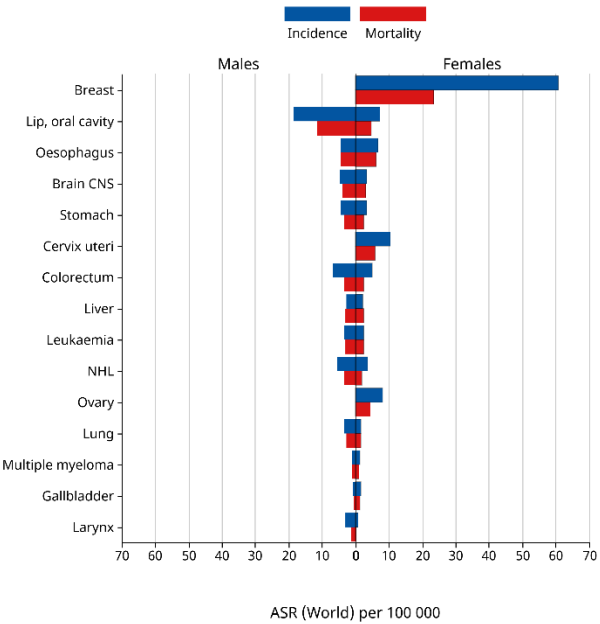


ASR (World) per 100 000

Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025



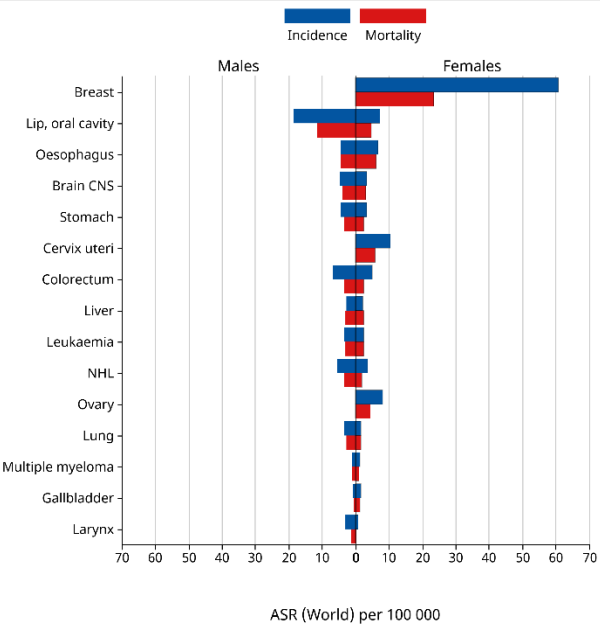
**Age-Standardized Rate (World) per 100 000, Incidence and Mortality, Males and Females, age [35-49], in 2022**  
Pakistan  
(Top 15 cancer sites)



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025



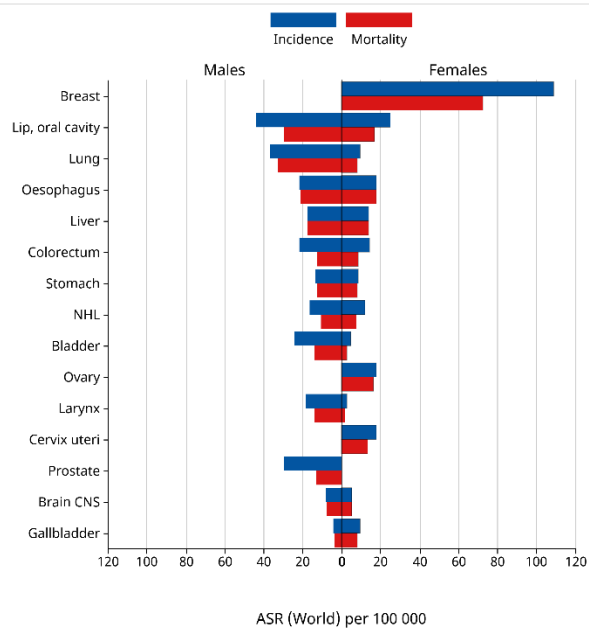
**Age-Standardized Rate (World) per 100 000, Incidence and Mortality, Males and Females, age [45-59], in 2022**  
Pakistan  
(Top 15 cancer sites)



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025



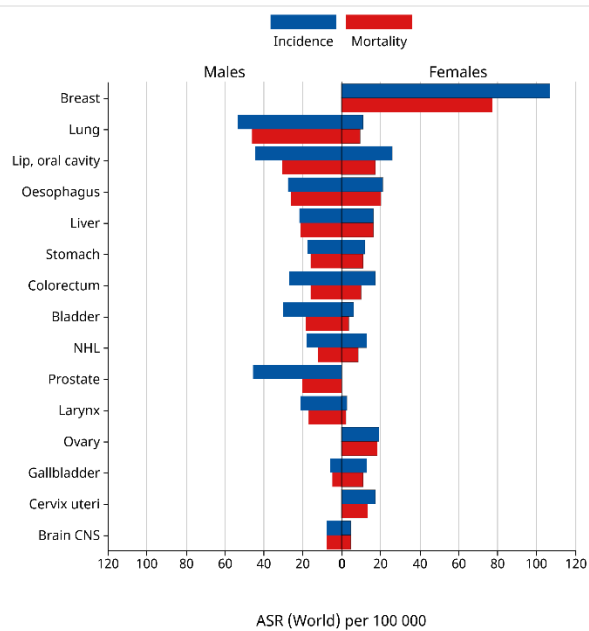
**Age-Standardized Rate (World) per 100 000, Incidence and Mortality, Males and Females, age [55-69], in 2022**  
 Pakistan  
 (Top 15 cancer sites)



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
 Data version : Globocan 2022 (version 1.1)  
 © All Rights Reserved 2025



**Age-Standardized Rate (World) per 100 000, Incidence and Mortality, Males and Females, age [60-84], in 2022**  
 Pakistan  
 (Top 15 cancer sites)



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
 Data version : Globocan 2022 (version 1.1)  
 © All Rights Reserved 2025



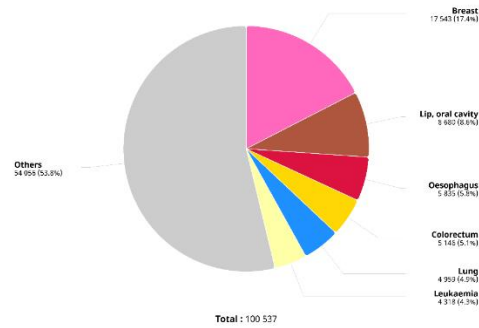


### Top 5 prevalent cancer types in Pakistan:

Rank	Cancer Type	% of Cases	Gender Trend	Source (Verified)
1	<b>Breast Cancer</b>	~38.8% of all female cancers	Mostly female	National Cancer Registry of Pakistan (JCPSP 2022)
2	<b>Oral Cancer</b>	~11.6% of all male cancers	Mostly male	NCRP (2015–2019 report)
3	<b>Colorectal Cancer</b>	~6.4% males, ~3.3% females	Both genders	NCRP (JCPSP)
4	<b>Liver Cancer</b>	~6.7% of male cancers	Mostly male	NCRP (JCPSP)
5	<b>Lung Cancer</b>	~6.0% of male cancers	Mostly male	NCRP (JCPSP)

## According to year prevalent

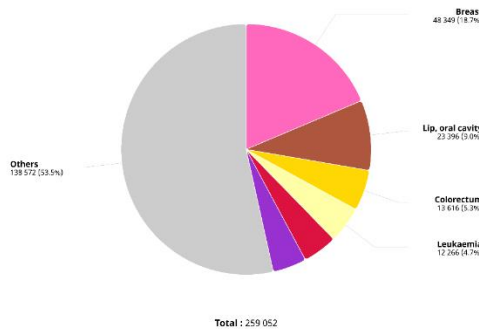
Estimated number of prevalent cases (1-year), Both sexes, in 2022  
Pakistan  
All cancers



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025



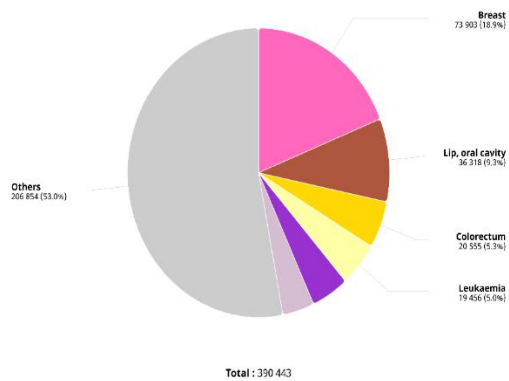
Estimated number of prevalent cases (3-year), Both sexes, in 2022  
Pakistan  
All cancers



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025



Estimated number of prevalent cases (5-year), Both sexes, in 2022  
Pakistan  
All cancers



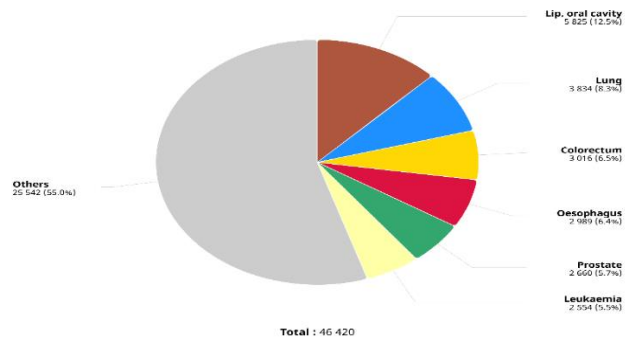
Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025



## According to Gender:

### Male

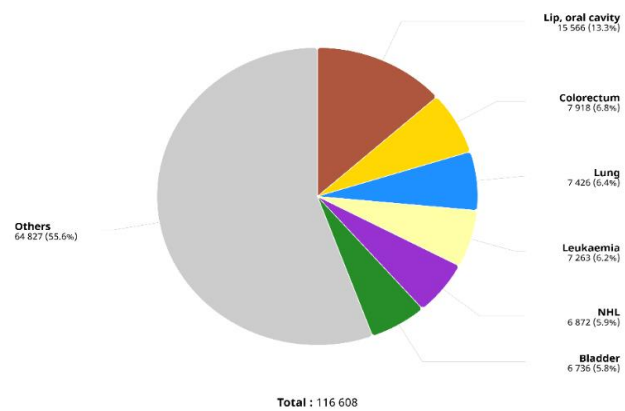
Estimated number of prevalent cases (1-year), Males, in 2022  
Pakistan  
All cancers



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

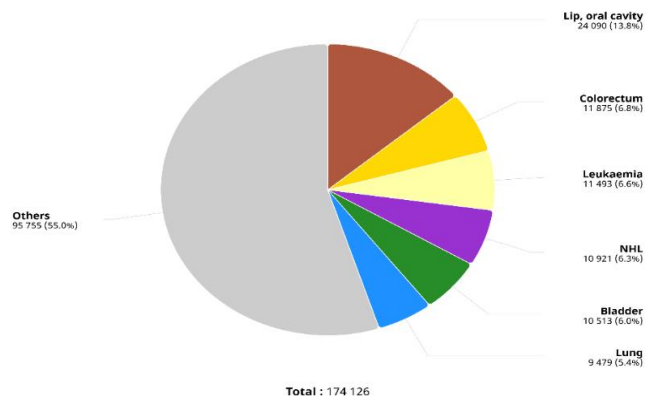
Estimated number of prevalent cases (3-year), Males, in 2022  
Pakistan  
All cancers



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

Estimated number of prevalent cases (5-year), Males, in 2022  
Pakistan  
All cancers



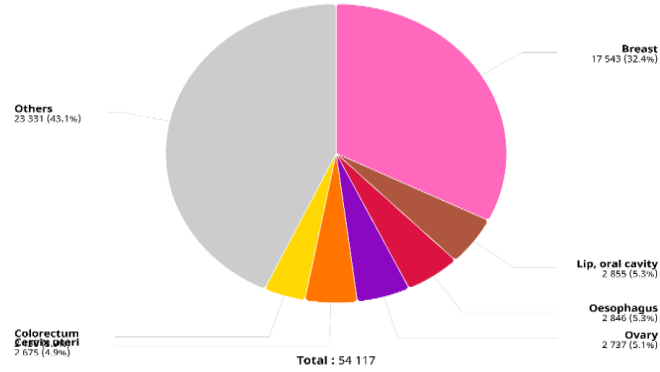
Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

## Female

### Estimated number of prevalent cases (1-year), Females, in 2022

Pakistan  
All cancers

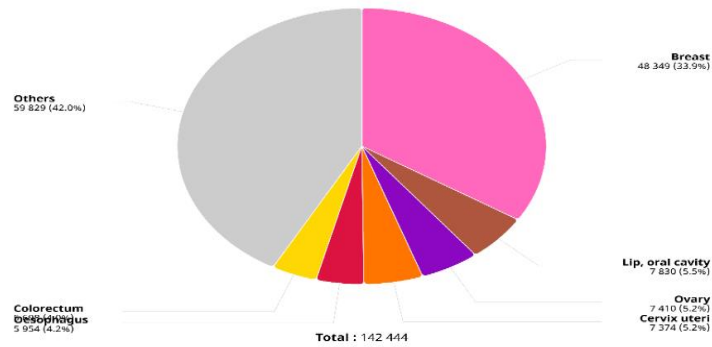


Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

### Estimated number of prevalent cases (3-year), Females, in 2022

Pakistan  
All cancers

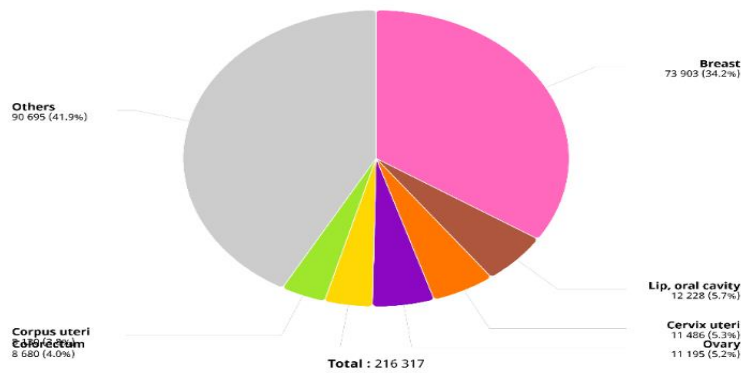


Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

### Estimated number of prevalent cases (5-year), Females, in 2022

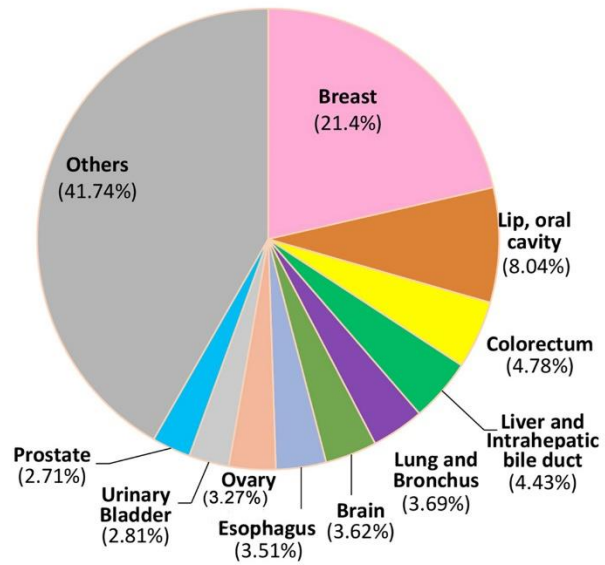
Pakistan  
All cancers



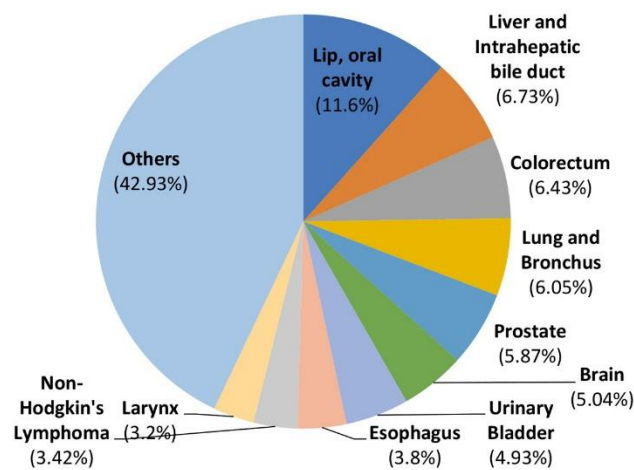
Cancer TODAY | IARC - <https://gco.iarc.who.int/today>  
Data version : Globocan 2022 (version 1.1)  
© All Rights Reserved 2025

International Agency  
for Research on Cancer  
World Health  
Organization

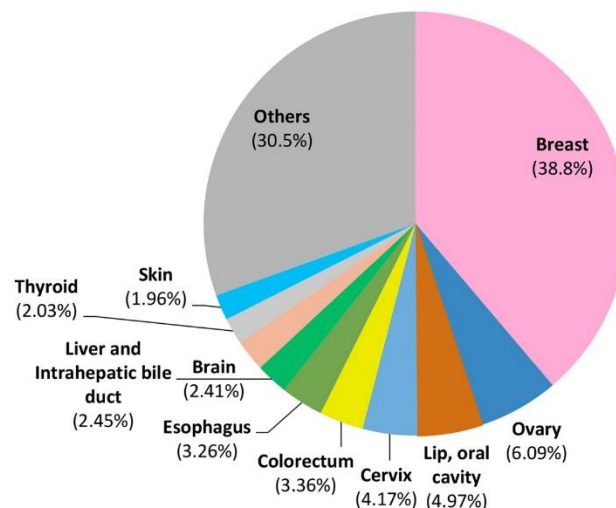
(a) Top ten cancers among both genders, all age groups (n=269,707)



(b) Top ten cancers among males, all age groups (n=124,795)



(c) Top ten cancers among females, all age groups (n=144,912)



## In Punjab Province:

### Punjab Cancer Registry Report-2024



### PUNJAB CANCER REGISTRY REPORT-2024

#### Background data

- In Lahore, the Punjab Cancer Registry (PCR) was established in 2005.
- The Shaukat Khanum Memorial Trust supports the Registry.
- There are 28 collaborating centers in the Registry.
- There are 55 members of the Registry, including an 8-member Governing Council.

#### The preliminary report for 2024

- The study's geographic area is the Lahore district in Punjab, Pakistan.
- The review period is January 1, 2024, to December 31, 2024.
- 6,986 new cases of cancer were reported overall.
- 3,977 (56.93%) females, 3,008 (43.06%) males, and 1 (0.01%) neuter.
- Children aged  $\leq 18$  years: 554 (7.93%) and adults: 6,432 (92.07%).
- This report only covers the top ten cancers. As a result, the sum is not 100%.
- This report is preliminary and will be revised in light of the latest information.

**Table 1. Total and frequently reported cancers, by sex and age group, Lahore, Pakistan, January 1, 2024 to December 31, 2024, the Punjab Cancer Registry.**

Total cases reported	Count	%
Total malignancies reported	<b>6,986</b>	<b>100</b>
Female	3,977	56.93
Male	3,008	43.06
Neuter	1	0.01
<u>Ten most frequently reported cancers: all age groups, both sexes combined (N=6,986)</u>		
Breast	1,629	23.3
Brain and nervous system	426	6.1
Liver and intrahepatic bile ducts	381	5.5
Colon, rectum, and anus	361	5.2
Non-Hodgkin lymphoma	299	4.3
Lip and oral cavity	281	4.0
Leukemia	274	3.9
Corpus uteri and uterus	255	3.7
Trachea, bronchus, and lung	229	3.3
Urinary bladder	213	3.1

Other reporting districts (24 in number) reported 17,372 cases to the Registry in 2024.

January 27, 2025.

Central Office at Shaukat Khanum Memorial Cancer Hospital and Research Center  
7-A, Block R-3, M.A. Johar Town, P.O. Box 13014, Postal Code 54782, Lahore, Pakistan. Ph. +92-42-35905000 Ext. 4243/4246.

## Punjab Cancer Registry Report-2023



### **PUNJAB CANCER REGISTRY REPORT-2023**

#### **Background information**

- The Punjab Cancer Registry (PCR) was established in 2005 in Lahore.
- The Shaukat Khanum Memorial Trust sponsors the Registry.
- The Registry has twenty-eight collaborating centers.
- The Registry has fifty-three members, nine of whom make up the Governing Council.

#### **The 2023 preliminary report**

- Geographic area under study: the Lahore district, Punjab, Pakistan.
- Review period: From January 1, 2023, to December 31, 2023 (1 year).
- Total: 6,271 new cancer diagnoses reported.
- Females: 3,489 (55.64%); males: 2,780 (44.33%); and neuters: 2 (0.03%).
- Adults: 5,803 (92.54%), and children: 468 (7.46%).
- This is a report on the top ten cancers only. Therefore, the total does not add up to 100%.
- This report will be updated when more data becomes available.

**Table 1. Total and frequently reported cancers, by sex and age group, in Lahore, Pakistan, from January 1, 2023, to December 31, 2023, the Punjab Cancer Registry.**

<b>Total cases reported</b>	<b>Count</b>	<b>%</b>
Total malignancies reported	<b>6,271</b>	<b>100</b>
Female	3,489	55.64
Male	2,780	44.33
Neuter	2	0.03
<b>Ten most frequently reported cancers: all age groups, both sexes combined (N=6,271)</b>		
Breast	1,391	22.2
Brain and nervous system	417	6.7
Colon, rectum, and anus	326	5.2
Lip and oral cavity	286	4.6
Non-Hodgkin lymphoma	282	4.5
Liver and intrahepatic bile ducts	261	4.2
Trachea, bronchus, and lung	235	3.8
Leukemia	234	3.7
Corpus uteri and uterus	217	3.5
Urinary bladder	200	3.2

August 20, 2024

Punjab Cancer Registry Report-2022



**PUNJAB CANCER REGISTRY REPORT-2022**

**Background information:**

- In Lahore, the Punjab Cancer Registry (PCR) was established in 2005.
- The Registry is sponsored by the Shaukat Khanum Memorial Trust.
- There are 27 centers that work with the Registry.
- There are 52 members of the Registry, including a Governing Council of 10.

**This is a preliminary report for 2022:**

- The Lahore district in Punjab, Pakistan, is the study's catchment area.
- The evaluation period is from January 1 to December 31, 2022 (1-year).
- There have been 6,661 new cancer diagnoses overall.
- Male: 2,792 (41.92%); female: 3,869 (58.08%).
- 6,156 adults (92.42%) and 505 children (under 18 years) made up the patient population.
- This analysis only covers the top ten cancers. As a result, the sum does not equal 100%.
- This report is merely preliminary; it will be updated as new information becomes available.

**Table 1. The table shows the most common malignancies by sex and age group reported in the Pakistani district of Lahore from January 1 to December 31, 2022, according to the Punjab Cancer Registry.**

<b>Total cases reported</b>	<b>Count</b>	<b>%</b>
Total malignancies reported	<b>6,661</b>	<b>100</b>
Female	3,869	58.08
Male	2,792	41.92
<b>Ten most frequently reported cancers in all age groups, both sexes combined (N=6,661)</b>		
Breast	1,659	24.9
Brain and nervous system	425	6.4
Colon, rectum, and anus	327	4.9
Non-Hodgkin lymphoma	327	4.9
Lip and oral cavity	278	4.2
Leukemia	257	3.9
Urinary bladder	243	3.7
Corpus uteri and uterus	242	3.6
Prostate	218	3.3
Trachea, bronchus, and lung	183	2.7

Central Office at the Shaukat Khanum Memorial Cancer Hospital and Research Center  
7-A, Block R-3, M.A. Johar Town, P O Box 13014, Postal Code 54782, Lahore, Pakistan. Ph. +92-42-35905000 Ext. 4243/4246.  
Email: [pcr@skm.org.pk](mailto:pcr@skm.org.pk)



### Punjab Cancer Registry Report-2022

#### Ten most frequently reported cancers in adult females (N=3,652)

Breast	1,632	44.7
Corpus uteri and uterus	242	6.6
Colon, rectum, and anus	130	3.6
Non-Hodgkin lymphoma	129	3.5
Ovary	128	3.5
Brain and nervous system	115	3.2
Lip and oral cavity	97	2.7
Cervix	87	2.4
Liver and intrahepatic bile ducts	67	1.8
Thyroid	59	1.6

#### Ten most frequently reported cancers in adult males (N=2,504)

Brain and nervous system	220	8.8
Prostate	218	8.7
Urinary bladder	195	7.8
Colon, rectum, and anus	191	7.6
Lip and oral cavity	179	7.1
Non-Hodgkin lymphoma	168	6.7
Trachea, bronchus, and lung	133	5.3
Kidney	106	4.2
Liver and intrahepatic bile ducts	87	3.5
Skin	84	3.4

#### Ten most frequently reported cancers in adults (>18 years (N=6,156))

Breast	1,658	26.9
Brain and nervous system	335	5.4
Colon, rectum, and anus	321	5.2
Non-Hodgkin lymphoma	297	4.8
Lip and oral cavity	276	4.5
Corpus uteri and uterus	242	3.9
Urinary bladder	241	3.9
Prostate	218	3.5
Trachea, bronchus, and lung	183	3.0
Liver and intrahepatic bile ducts	154	2.5

#### Ten most frequently reported cancers in children (<=18 years (N=505))

Acute lymphoblastic leukemia	126	25.0
Glioma	82	16.2
Hodgkin lymphoma	37	7.3
Non-Hodgkin lymphoma	30	5.9
Ewing Sarcoma / PNET	28	5.5
Osteosarcoma	27	5.4
Acute myeloid leukemia	18	3.6
Germ cell tumor	17	3.4
Retinoblastoma	12	2.4
Nephroblastoma	9	1.8

February 27, 2023.

### Cancer Statistics Summary

Cancer Type	Global Rank	Pakistan Rank	Incidence Rate	Gender (M/F)	Most Affected Region	Source
Breast Cancer	1	1	12.5% (Global)	F	Punjab	WHO, GLOBOCAN, SKMCH
Lung Cancer	2	3	12.2% (Global)	M	Sindh	WHO, GLOBOCAN
Colorectal Cancer	3	4	10.7% (Global)	M/F	Punjab	WHO, GLOBOCAN
Prostate Cancer	4	-	7.3% (Global)	M	-	WHO
Oral Cancer	9	2	3.9% (Global)	M	Sindh	Shaukat Khanum Registry

## Part B – Research Direction Identification

### Selected Cancer: Oral Cancer Diagnosis

Oral cancer was chosen due to:

1. High prevalence in Pakistan, particularly among males
2. Potential for **early detection using non-invasive imaging**
3. Availability of open datasets for AI modeling
4. Opportunity to develop **mobile or computer-aided screening tools**

#### AI Research Directions in Early Detection

Research Area	Description	AI Approach / Tools
Early lesion detection	Detect oral lesions in clinical images before malignancy	CNN, Vision Transformers, Mobile AI apps
Histopathology analysis	Classify tissue as normal, precancerous, or malignant	ResNet, DenseNet, U-Net, segmentation
Multi-modal modeling	Combine images + patient demographics	Ensemble learning, multi-input neural networks
Explainable AI	Provide interpretable results for clinicians	Grad-CAM, LIME, SHAP
Risk prediction	Predict cancer progression or recurrence	Random Forest, SVM, Gradient Boosting

**Challenges:** Limited datasets, variability in image quality, ethical concerns, low-resource deployment.

**Opportunities:** Early screening apps, multi-modal models, explainable AI integration.

## Part C – Dataset Exploration

This portion is still pending, furthermore exploration will be continue.

Dataset	Type	Size	Access
<b>Oral Cancer Dataset (Kaggle)</b>	Clinical photographs (benign/malignant)	~1,200 images	Kaggle - Oral Cancer
<b>ISBI Oral Lesion Dataset (NIH)</b>	Annotated oral mucosal lesion images	2,000+	NIH ISBI Challenge Dataset
<b>AIIMS Oral Cancer Dataset (India)</b>	Real hospital images	~4,000	Need institutional access
<b>Augmented Oral Cancer Dataset (GitHub)</b>	Augmented versions for CNN training	Variable	GitHub search: <i>Oral Cancer Classification Dataset</i>

**Some more dataset references are:**

<b>Dataset Name (Year)</b>	<b>Type (Images / Tabular / Genomic)</b>	<b>Source / Link</b>	<b>Description</b>	<b>Size</b>
NDB-UFES Oral Cancer Dataset (2023)	Images + Tabular	<a href="https://data.mendeley.com/datasets/bbmmm4wgr8">https://data.mendeley.com/datasets/bbmmm4wgr8</a>	Contains 237 histopathologic images of oral squamous cell carcinoma and leukoplakia, plus sociodemographic and clinical data from UFES University Hospital (Brazil).	237 images + tabular metadata
Histopathological Imaging Database for Oral Cancer Analysis (2022)	Images (H&E Histopathology)	<a href="https://data.mendeley.com/datasets/ftmp4cvtmb">https://data.mendeley.com/datasets/ftmp4cvtmb</a>	Digitized H&E-stained microscopic images of oral squamous cell carcinoma and normal epithelium at 100× and 400× magnifications.	1,224 images (2 magnifications)
ORCHID: Oral Cancer Histology Image Database (2024)	Images (Histology)	<a href="https://www.nature.com/articles/s41597-024-03836-6">https://www.nature.com/articles/s41597-024-03836-6</a>	Large-scale histology dataset of oral cancer and precancerous lesions for AI-based diagnosis and segmentation with expert annotations	>3,000 high-resolution tiles
MODID – Multispectral Oral Disease Image Dataset (2022)	Images (Multispectral)	<a href="https://doi.org/10.5061/dryad.nvx0k6dxw">https://doi.org/10.5061/dryad.nvx0k6dxw</a>	Spectral images of oral tissues (healthy, precancerous, cancerous) captured in 16 bands (460–600 nm).	~480 spectral image sets
OCDC – Oral Squamous Cell Carcinoma Dataset (2023)	Images (H&E Histology + Masks)	<a href="https://arxiv.org/abs/2303.10172">https://arxiv.org/abs/2303.10172</a>	H&E-stained histology slides of OSCC with pixel-level tumor annotations for segmentation research.	~1,020 images (640×640 px)

## Conclusion

Oral cancer is a **high-impact research area** for AI due to its prevalence in Pakistan and the feasibility of using clinical and histopathology images. AI models can aid in **early detection, lesion classification, risk prediction, and mobile screening**, bridging gaps in healthcare access and improving patient outcomes.

## References

1. World Health Organization. *Cancer Fact Sheet*. <https://www.who.int/news-room/fact-sheets/detail/cancer>
2. GLOBOCAN 2022, *Global Cancer Observatory*. <https://gco.iarc.fr/>
3. Shaukat Khanum Cancer Registry Reports, 2022. <https://shaukatkhanum.org.pk>
4. National Cancer Registry of Pakistan (NCRP), JCPSP 2015–2019
5. Kaggle Oral Cancer Image Dataset. <https://www.kaggle.com/datasets>
6. ISBI Oral Lesion Histopathology Dataset. <https://isbi.org>
7. Review papers: “Artificial Intelligence in Oral Cancer” – MDPI Diagnostics 2023