Replication study of Higgs Boson Search in the Four-Lepton Final State Using ATLAS Open Data*

Hayato Isa
University of California, Berkeley†
(Dated: December 12, 2025)

Abstract: This paper presents a replication study of the Higgs boson search in the four-lepton final state ($H \to ZZ^* \to 4\ell$) using 2016 ATLAS Open Data. The analysis follows the key methodological steps of the original 2012 ATLAS and CMS discovery analyses, including event selection, reconstruction of the four-lepton invariant mass (m_{4l}), and comparison with Monte Carlo simulations. Tight selection criteria are applied to suppress backgrounds, targeting events with exactly four leptons consistent with detector acceptance, momentum thresholds, isolation, and vertex requirements. The m_{4l} distribution is constructed from the selected data and simulated signal and background, revealing a peak consistent with a Higgs boson of mass $m_H \approx 125$GeV. The maximum likelihood estimate for the signal strength is $\hat{\mu} = 1.25$. Using the profile likelihood ratio as a test statistic, the analysis yields an observed significance of 2.87σ ($p \approx 2.05 \times 10^{-3}$), providing moderate statistical evidence for the existence of the Higgs boson as predicted by the Standard Model and supporting the mechanism of electroweak symmetry breaking.

Keywords: Higgs boson; Four-lepton decay; ATLAS Open Data; Standard Model

CONTENTS

I. INTRODUCTION

In gauge theories, gauge bosons are massless unless the symmetry is broken, but the W and Z bosons are observed to have nonzero mass. This apparent discrepancy is resolved by introducing the Higgs field, which allows gauge bosons to acquire mass via spontaneous symmetry breaking and simultaneously predicts the existence of the Higgs boson [1]. For many years, the Higgs field and the associated Higgs boson remained theoretical constructs lacking experimental confirmation. This situation changed in 2012, when the ATLAS and CMS experiments at the Large Hadron Collider (LHC) provided definitive experimental evidence for the Higgs boson, and validating the mechanism of mass generation in the Standard Model [2–5].

Among the various Higgs decay modes, the four-lepton final state ($H \to ZZ^* \to 4\ell$) is particularly advantageous for Higgs searches. This channel benefits from the high-precision measurement of charged leptons, enabling the invariant mass of the four-lepton system (m_{4l}) to be reconstructed with excellent resolution. Although several background processes can produce four-lepton candidates, such as

$$\begin{cases} qq \to Z + \text{jets} \\ qq \to ZW/Z\gamma \\ gg/qq \to tt \end{cases}$$

, these are all significantly reducible. In addition, the primary irreducible backgrounds

$$\begin{cases} q\bar{q} \to ZZ^* \to 4\ell \\ gg \to ZZ^* \to 4\ell \end{cases}$$

do not involve a resonant intermediate state, resulting in a smooth m_{4l} distribution without fake peaks. Consequently, if the Higgs boson exists, a signal peak corresponding to the Higgs boson can be distinguished above the previously established lower mass bound of 114.4 GeV [6].

In this paper, we reproduce the 2012 Higgs discovery analysis by focusing on the four-lepton channel. The paper follows the key methodological steps of the original ATLAS and CMS analyses, including event selection, invariant mass reconstruction, and comparison with Monte Carlo simulations[4, 5]. Finally, we use the profile likelihood ratio to compute the observed significance [7], and provide statistical evidence supporting the Higgs mechanism.

* The complete analysis code is publicly available at `https://github.com/Hayathorium/higgs-4l-opendata`.
† iamhayatoisa@berkeley.edu

A. Higgs Boson Production Modes and the $H \to ZZ^* \to 4\ell$ Channel

Within the framework of the Standard Model, the primary Higgs boson production mechanisms include gluon–gluon fusion (ggF), vector boson fusion (VBF), associated production with vector bosons (WH and ZH), associated production with top–antitop or bottom–antibottom pairs (ttH and bbH), and single-top associated production (tH)[8]. The corresponding theoretical predictions for the production cross sections are summarized in Table I.

TABLE I. Standard Model Higgs boson production cross sections at $\sqrt{s} = 13$ TeV for $m_H = 125$ GeV [9]

Process	**Cross Section σ (pb)**
$gg \to H$ (a)	48.58
$qq \to qqH$ (b)	3.782
$qq \to WH$ (c)	1.373
$qq \to ZH$ (c)	0.8839
$gg \to ttH$ (d)	0.5071
$gg \to bbH$ (d)	0.4880
$bq \to tqH$ (e)	0.07425
Total σ (all-in)	**55.71**

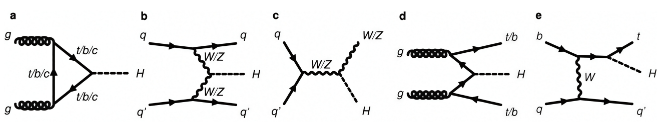

FIG. 1. Feynman diagrams for single Higgs boson production modes in the Standard Model [10]

For the $H \to ZZ^* \to 4l$ decay channel, assuming a Higgs boson mass of $m_H = 125$GeV, the Standard Model predicts the following branching ratios[9][12]:

$$BR(H \to ZZ^*) = (2.619 \pm 0.026)\%$$
$$BR(Z \to e^+e^-) = (3.3632 \pm 0.0042)\%$$
$$BR(Z \to \mu^+\mu^-) = (3.3662 \pm 0.0066)\%$$

The resulting branching ratio for the four-lepton final state is therefore

$$\begin{aligned} &BR(H \to ZZ^* \to 4l) \\ &= BR(H \to ZZ^*)\left[BR(Z \to e^+e^-) + BR(Z \to \mu^+\mu^-)\right]^2 \\ &= (1.186 \pm 0.012) \times 10^{-4} \end{aligned}$$

Although the branching ratio for $H \to ZZ^* \to 4\ell$ is relatively small, this channel remains particularly advantageous for Higgs boson studies for several reasons, as discussed in the following section.

II. ANALYSIS METHODOLOGY

The $H \to ZZ^* \to 4\ell$ decay modes are considered as the cleanest modes for Higgs boson studies. For a Higgs mass $m_H < 2m_Z$, the Higgs decays into one on-shell Z boson and one off-shell Z^* boson. Each Z or Z^* subsequently decays into a pair of electrons or muons, as illustrated in Figure II, with the leptons originating from the same vertex sharing the same flavor. The resulting final states are therefore four charged leptons ($e^+e^-e^+e^-$, $e^+e^-\mu^+\mu^-$, or $\mu^+\mu^-\mu^+\mu^-$). Tau leptons are generally not considered because they decay rapidly into lighter leptons and neutrinos before reaching the detector.

Due to the excellent momentum resolution for electrons and muons, the invariant mass of the four-lepton system (m_{4l}) can be reconstructed with high precision, producing a sharp peak at m_H in the presence of a Higgs signal. These features allow the application of strict event selection criteria, enabling the Higgs signal to be identified with high significance. For these reasons, despite its very small branching ratio, the $H \to ZZ^* \to 4\ell$ decay channel is widely called "Golden Channel" for Higgs boson detection.

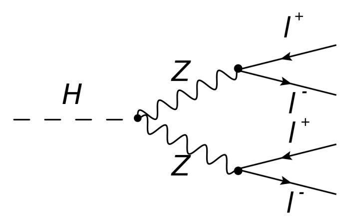

FIG. 2. Feynman diagram for the Higgs boson decay into four charged leptons [13]

For this analysis, we utilize the dataset II, together with the variables listed in Table III and follow the methodology outlined in [4].

To scale the Monte Carlo simulation events to the integrated luminosity of the experimental dataset ($\mathcal{L}_{int} \approx 10000pb^{-1}$) and to account for detector and reconstruction inefficiencies, each event is assigned a weight w:

$$w = \underbrace{\frac{mcWeight}{SumWeights}}_{\text{Normalization by sample}} \underbrace{XSection \cdot \mathcal{L}_{int}}_{\text{Expected \#events}} \underbrace{\prod scaleFactor}_{\text{Efficiency Correction}}$$

Event selection begins with loading the full dataset 1 and selecting events containing exactly four leptons, consistent with the $H \to ZZ^* \to 4l$ channel 2. Detector coverage restricts electrons and muons to the pseudorapidity ranges $|\eta| < 2.47$ and $|\eta| < 2.7$, respectively 3[14]. Leptons in each event are ordered in descending transverse momentum, and events are required to satisfy $p_{T,l_1} > 20$GeV, $p_{T,l_2} > 15$GeV, $p_{T,l_3} > 10$GeV, and $p_{T,l_4} > 7$GeV 4. This selection reduces misidentified leptons originating from jets, charged hadrons, or heavy-quark decays and also improves the accuracy of the reconstructed four-lepton invariant mass (m_{4l}).

TABLE II. Summary of data and Monte Carlo samples used in the analysis [11]

File Name	Description
Real Data Files	
data_A.4lep.root	ATLAS Run A data containing four-lepton candidate events.
data_B.4lep.root	ATLAS Run B data containing four-lepton candidate events.
data_C.4lep.root	ATLAS Run C data containing four-lepton candidate events.
data_D.4lep.root	ATLAS Run D data containing four-lepton candidate events.
Signal Monte Carlo Samples	
mc_341947.ZH125_ZZ4lep.4lep.root	Associated production (ZH, $m_H = 125$ GeV).
mc_341964.WH125_ZZ4lep.4lep.root	Associated production (WH, $m_H = 125$ GeV).
mc_344235.VBFH125_ZZ4lep.4lep.root	Vector Boson Fusion (VBF, $m_H = 125$ GeV).
mc_345060.ggH125_ZZ4lep.4lep.root	Gluon–gluon fusion (ggF, $m_H = 125$ GeV).
Background Monte Carlo Samples	
mc_363490.llll.4lep.root	Irreducible $ZZ^* \to 4\ell$ background.
mc_410000.ttbar_lep.4lep.root	$t\bar{t}$ background (leptonic).
mc_361108.Ztautau.4lep.root	$Z \to \tau\tau$ background ($Z +$ jets).
mc_361107.Zmumu.4lep.root	$Z \to \mu\mu$ background ($Z +$ jets).
mc_361106.Zee.4lep.root	$Z \to ee$ background ($Z +$ jets).

TABLE III. Descriptions of selected ATLAS mini-tree variables

Variable	Description
lep_n	Total number of reconstructed leptons in the event.
lep_pt	Transverse momentum of each lepton, p_T (MeV).
lep_eta	Pseudorapidity of each lepton, $\eta = -\ln\tan(\theta/2)$.
lep_phi	Azimuthal angle of each lepton in the transverse plane (radians).
lep_E	Energy of each lepton in GeV.
lep_isTightID	Boolean flag indicating whether the lepton passes the tight identification criteria.
lep_type	Lepton particle type code: 11 for electron, 13 for muon.
lep_charge	Electric charge of the lepton (+1 or -1).
lep_ptcone30	Sum of transverse momenta of tracks within $\Delta R = 0.3$ around the lepton, excluding itself.
lep_z0	Longitudinal impact parameter of the lepton relative to the primary vertex (mm).
lep_trackd0pvunbiased	Transverse impact parameter of the lepton relative to the primary vertex (mm).
mcWeight	Weight assigned to the event by the Monte Carlo generator, used for normalization.
XSection	Generator-level production cross-section of the simulated process (pb).
SumWeights	Sum of event generator weights, used for overall normalization of the sample.
scaleFactor_PILEUP	Correction factor to account for differences in pileup between data and simulation.
scaleFactor_ELE	Electron identification and reconstruction efficiency correction factor.
scaleFactor_MUON	Muon identification and reconstruction efficiency correction factor.
scaleFactor_PHOTON	Photon identification and reconstruction efficiency correction factor.
scaleFactor_TAU	Tau lepton identification efficiency correction factor.
scaleFactor_BTAG	b-jet tagging efficiency correction factor.
scaleFactor_LepTRIGGER	Efficiency correction for the lepton trigger selection.
scaleFactor_PhotonTRIGGER	Efficiency correction for the photon trigger selection.

Leptons are further required to be spatially separated by $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} > 0.1$ for same-flavor pairs and $\Delta R > 0.2$ for different-flavor pairs 5, preventing mis-reconstruction due to overlapping tracks. Leptons satisfying the ATLAS "tight" identification criteria [15] are selected 6. Since the Z boson decays almost instantaneously ($\tau_Z \sim 3 \times 10^{-25}$s), the decay products are expected to originate near the primary vertex. Thus, leptons are required to satisfy a longitudinal impact parameter cut of $z_0 < 10$mm 7. Additionally, to suppress cosmic-ray muon contamination, muons are required to have a small transverse impact parameter $z_d < 1$mm 8.

The analysis also exploits lepton isolation. Higgs decays produce isolated leptons, while leptons from background processes often arise within jets. Events are therefore required to satisfy $\sum p_T^{cone}/p_{T,l} < 0.15$ 9. For reconstruction of Z candidates, the lepton pair whose invariant mass (m_{12}) is closest to the Z boson mass must satisfy 50GeV $< m_{12} <$ 106GeV. The remaining lepton pair, with invariant mass (m_{34}), must satisfy $m_{min} < m_{34} <$ 115GeV, where m_{min} varies monotonically from 17.5GeV at $m_{4l} =$ 120GeV to 50GeV at

$m_{4l} = 190\text{GeV}$ 10[16].

Following event selection, we construct a stacked histogram of m_{4l} for both the Higgs signal and the background, weighted by w, and overlay the corresponding distribution from observed experimental data. Fine binning (5 GeV) is chosen to enhance sensitivity. Finally, the binned m_{4l} distribution is used to compute the maximum likelihood estimate of the signal strength parameter and the observed significance of the Higgs signal.

Systematic uncertainties can affect both the signal strength estimate $\hat{\mu}$ and the observed significance. The main sources of systematic uncertainty are integrated luminosity, theoretical cross-section and modeling uncertainties, and lepton reconstruction and identification efficiencies. In this analysis, we neglect systematic uncertainties to focus on reproducing the statistical methodology of the original study.

III. RESULT

The event selection procedure results are summarized in the cutflow presented in Table IV, with all entries normalized to an integrated luminosity of 10000pb^{-1}. The corresponding distribution of the four-lepton invariant mass (m_{4l}) after selection is shown in Figure 3. The statistical uncertainty on the observed data is estimated as $\sigma_i = \sqrt{\#\text{counts}_i}$ for each bin i, based on the approximation that the number of events in each bin follows $Poisson(\#\text{counts}_i)$. The observed counts are generally consistent with the simulation predictions within these statistical uncertainties, indicating good agreement with the Standard Model expectation for a Higgs boson with $m_H = 125\text{GeV}$. Given an expected background of $\lambda = 1.5$ events in the [120,125]GeV bin (from background MC), if the true Higgs boson mass were significantly different from 125GeV, the probability of observing six events in the [120,125]GeV mass bin would be

$$P(N_{[120,125]} = 6) = \frac{e^{-\lambda}\lambda^6}{6!} \approx 0.0035$$

which is very small. Therefore, the presence of a peak in this mass region supports the hypothesis $m_H \approx 125\text{GeV}$.

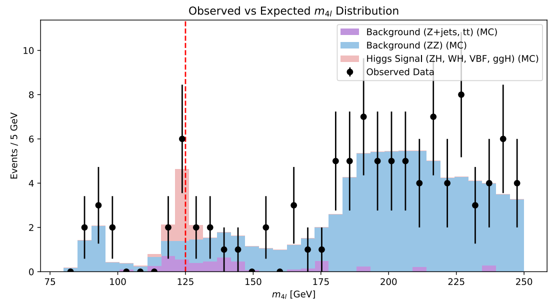

FIG. 3. Distribution of the reconstructed four-lepton invariant mass (m_{4l}) after event selection.
(Normalized to $\mathcal{L}_{int} = 10\text{fb}^{-1}$)

A. Significance

The analysis uses the profile likelihood ratio as the test statistic [7]. Let μ denote the signal strength parameter, s_i the expected Higgs signal, b_i the expected background, and k_i the observed number of events in bin i. Assuming $k_i \sim Poisson(\mu s_i + b_i)$, the likelihood function is given by

$$L(\mu) = \prod_i \frac{(\mu s_i + b_i)^{k_i} e^{-(\mu s_i + b_i)}}{k_i!}$$

The hypothesis test is formulated as

$$\begin{cases} H_0 : \mu = 0 \text{ (no Higgs signal)} \\ H_1 : \mu > 0 \text{ (Higgs signal present)} \end{cases}$$

Using the maximum likelihood estimate $\hat{\mu}$, the test statistic significance is expressed as

$$\sqrt{q_\mu} = \sqrt{-2\ln\frac{L(\mu = 0)}{L(\mu = \hat{\mu})}}$$

Prior to examining the observed data, an Asimov dataset ($k_i = s_i + b_i$) and $\hat{\mu} = 1$ are used to estimate the expected significance, yielding 2.30σ. This indicates that the event and bin selection criteria are robust for the Higgs search. Using the experimental data and the Monte Carlo predictions, the estimated signal strength is $\hat{\mu} = \arg\max_\mu L(\mu|s_i, b_i) = 1.25 \approx 1$, with an observed significance of 2.87σ ($p = 1 - \Phi(2.87) \approx 2.05 \times 10^{-3}$). These results provide moderate evidence for the existence of the Higgs boson as predicted by the Standard Model.

IV. SUMMARY AND CONCLUSIONS

The objective of this project was to reproduce the Higgs boson search in the four-lepton final state and to evaluate the observed significance of the Higgs signal,

TABLE IV. Number of events passing each selection step for background MC, signal MC, and observed data. ($\mathcal{L}_{int} = 10\text{fb}^{-1}$)

Selection Cut	**Background (ZZ)**	**Background (Z+jets, tt)**	**Signal**	**Data**
Cut 1: All events (preselection)	278.89	395.04	9.95	832
Cut 2: Exactly 4 leptons	277.54	392.09	9.84	827
Cut 3: Lepton $\|\eta\|$ acceptance	277.54	392.09	9.84	827
Cut 4: Lepton p_T thresholds cut	263.73	261.11	9.54	679
Cut 5: Lepton isolation (ΔR) cut	263.58	247.54	9.54	656
Cut 6: Tight lepton ID cut	184.17	100.46	6.48	311
Cut 7: Lepton longitudinal impact (z_0) cut	183.52	40.56	6.47	286
Cut 8: Muon transverse impact parameter (d_0) cut	183.52	40.09	6.47	285
Cut 9: Lepton isolation (p_T^{cone}) cut	171.17	11.84	5.71	225
Cut 10: Z candidate selection	132.73	5.07	4.86	166

providing experimental confirmation of the Higgs mechanism. To this end, we applied the event selection criteria to ATLAS open data and corresponding Monte Carlo simulations, following the methodology outlined in the original study [4]. The signal strength parameter $\hat{\mu}$ was calculated via the maximum likelihood estimation, and the observed significance was evaluated using the profile likelihood ratio.

The event selection targeted events containing exactly four reconstructed leptons with pseudorapidity, transverse momentum, and angular separation consistent with detector acceptance and resolution. Selected leptons were required to satisfy the ATLAS Tight identification criteria [15], originate near the primary vertex, and be isolated according to their p_T^{cone} values. Furthermore, events were required to contain opposite-sign, same-flavor lepton pairs with invariant masses consistent with the Z boson mass.

Following these selections, the binned four-lepton invariant mass (m_{4l}) distribution was constructed with 5GeV bins. The majority of observed counts lie within the estimated statistical uncertainties $\sqrt{\#\text{counts}_i}$, indicating satisfactory agreement between the data and simulation with $m_H = 125\text{GeV}$. The analysis yielded $\hat{\mu} = 1.25$ with an observed significance of 2.87σ, providing moderate evidence for the existence of the Higgs boson ($m_H = 125\text{GeV}$) in accordance with Standard Model predictions.

A. Reflects as a Student

Throughout this project, I gained hands-on experience with essential techniques in high-energy physics analysis, including event selection for background suppression and the evaluation of statistical significance. These skills provide a foundational framework for conducting more advanced research in the future.

[1] P. W. Higgs, Broken symmetries and the masses of gauge bosons, Phys. Rev. Lett. **13**, 508 (1964).

[2] A. Collaboration, A particle consistent with the higgs boson observed with the atlas detector at the large hadron collider, Science **338**, 1576 (2012).

[3] C. Collaboration, A new boson with a mass of 125 gev observed with the cms experiment at the large hadron collider, Science **338**, 1569 (2012).

[4] A. Collaboration, Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc, Phys. Lett. B **716**, 1 (2012).

[5] C. Collaboration, Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc, Phys. Lett. B **716**, 30 (2012).

[6] L. W. G. for Higgs boson searches, Search for the standard model higgs boson at lep, Phys. Lett. B **565**, 61 (2003).

[7] G. Cowan, K. Cranmer, E. Gross, and E. Vitells, Asymptotic formulae for likelihood-based tests of new physics, Eur. Phys. J. C **71**, 1554 (2011), arXiv:1007.1727.

[8] P. D. Group, Status of higgs boson physics (2023), review of Particle Physics, Higgs Boson Section.

[9] LHC Higgs Cross Section Working Group, Handbook of lhc higgs cross sections: 4. deciphering the nature of the higgs sector (2023), cERN-2017-002-M (LHCHWG-2016-01).

[10] A. Collaboration and C. Collaboration, Combined search for the standard model higgs boson decaying into a $b\bar{b}$ pair with the atlas and cms experiments, Nature **607**, 51 (2022).

[11] A. Collaboration, ATLAS open data: $Z \to \ell\ell$ candidate events, $\sqrt{s} = 8$ TeV (2017), cERN Open Data Portal.

[12] P. D. Group, Z boson (2024), review of Particle Physics, Z Boson Section.

[13] W. Commons, Feynman diagram for higgs to four leptons (2013), credit reference for Figure 2.

[14] A. Collaboration, The atlas inner detector: Expected performance (2010), aTL-PHYS-PUB-2010-013.

[15] A. Collaboration, Tools for estimating fake/non-prompt lepton backgrounds with the atlas detector at the lhc, JINST **18**, T11004, arXiv:2211.16178.

[16] A. Collaboration, Search for the higgs boson in the $h \to zz^* \to 4\ell$ decay channel in pp collisions at $\sqrt{s} = 8$ tev with the atlas detector (2012), aTLAS-CONF-2012-092.